

AD-A152 127

THE DEVELOPMENT OF AN ENLISTMENT STANDARDS MODEL FOR
THE NAVY AVIATION MACHINIST'S MATE (AD) RATING(U) NAVAL
POSTGRADUATE SCHOOL MONTEREY CA D A OSLUND ET AL

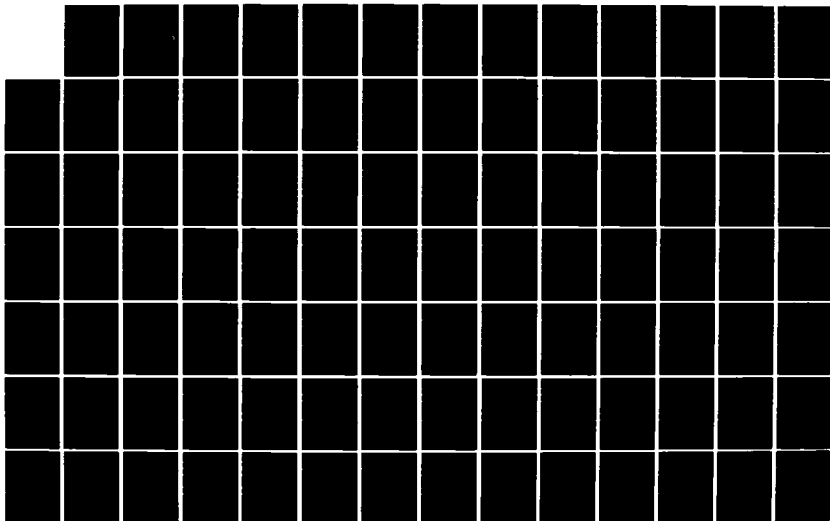
1/ 2

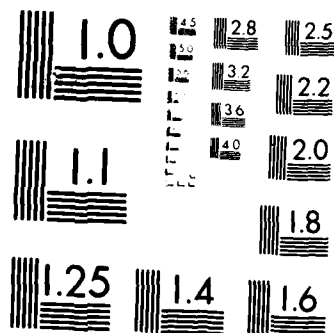
UNCLASSIFIED

JUN 84

F/G 5/9

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A152 127

(2)

NAVAL POSTGRADUATE SCHOOL

Monterey, California



DTIC
ELECTE
APR 5 1985
S B D

THESIS

DTIC FILE COPY

THE DEVELOPMENT OF AN
ENLISTMENT STANDARDS MODEL
FOR THE NAVY
AVIATION MACHINIST'S MATE (AD) RATING

by

Dwayne A. Oslund

and

J. S. A. Clark

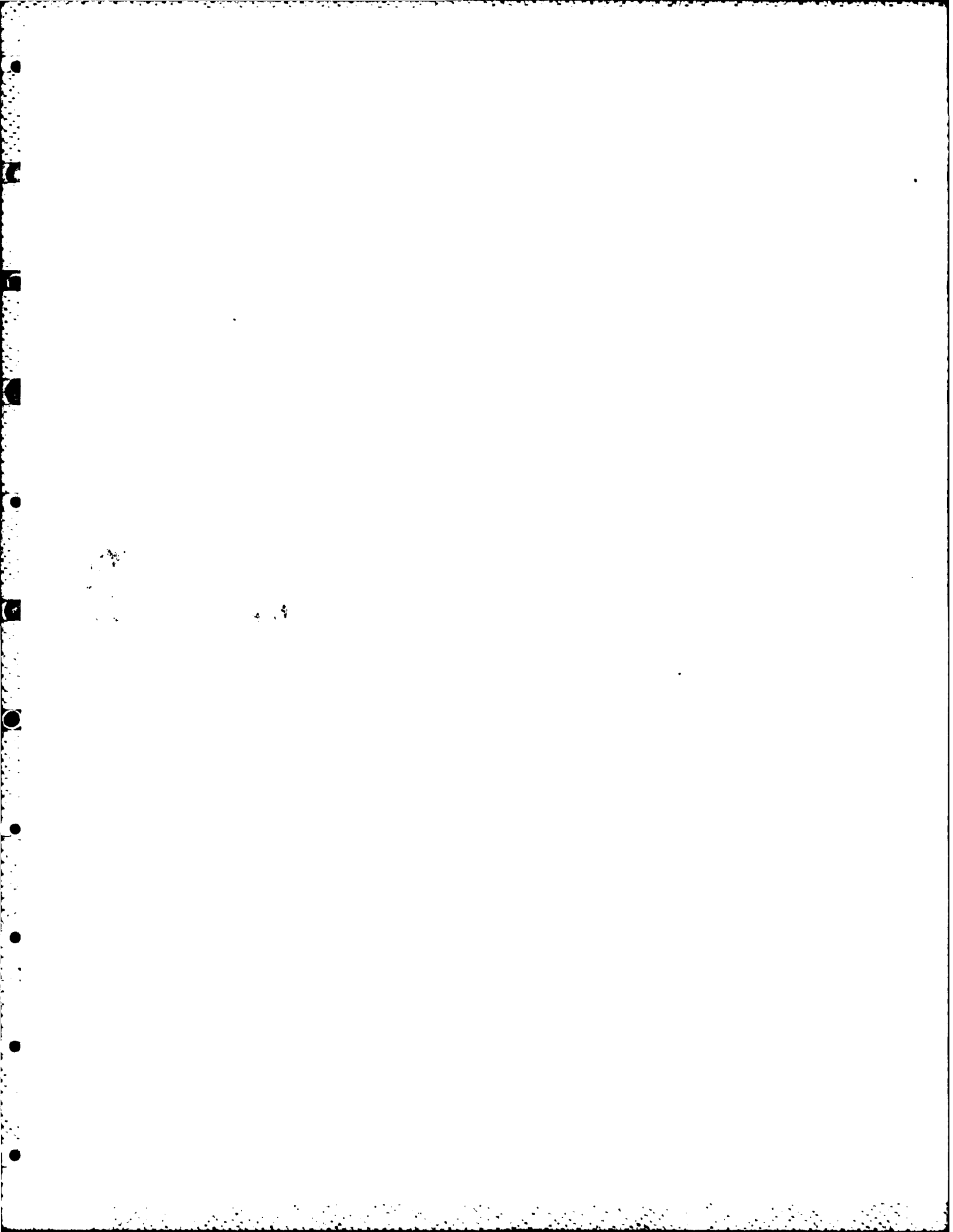
June 1984

Thesis Advisor:

W. E. McFarvey

Approved for public release, distribution unlimited.

85 03 18 016



Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
	AD-A152 127	
4. TITLE (and Subtitle) The Development of an Enlistment Standards Model for the Navy Aviation Machinist's Mate (AD) Rating		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis June 1984
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dwayne A. Oslund and J. S. A. Clark		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943		12. REPORT DATE June 1984
		13. NUMBER OF PAGES 113
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release, distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue in reverse side if necessary and identify by block number) Enlistment Standards, Enlistment Models, Regression Analysis, Discriminant Analysis, Utility Analysis, Model Validation, Selection Standards, Personnel Selection		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The purpose of this study is to present analytic techniques for developing enlistment standards models which attempt to improve upon existing methods. Alternative criteria for measuring successful operational performance, and a means of measuring utility are also provided. Another purpose of this study is to discover if the Navy's system of selecting personnel for the Aviation Machinist's Mate (AD) rating may be improved.		

DD FORM 1473

1 JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE

GPO 0102-LE-014-6601

Unclassified

1

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Two criteria were utilized in developing these models--length of service, and a composite measure of success that considers length of service, highest paygrade achieved, and reenlistment eligibility. Measures on individual's at the time of enlistment are used as predictor and discriminating variables. Six models are developed and analyzed using regression and discriminant techniques. Utility analysis is conducted on each of these models as a means for measuring their usefulness in monetary terms. Recommendations for future research are also presented.

S N 0102- LF-014-6601

Unclassified

2 SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Approved for public release; distribution unlimited.

The Development of an Enlistment Standards Model
for the
Navy Aviation Machinist's Mate (AD) Rating

by

Dwayne A. Oslund
Lieutenant Commander, United States Navy
B.S., Oregon State University, 1971

and

Captain J. S. A. Clark, RAE
Australian Army
B.E. (Civil), University of New South Wales, 1976

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
June 1984

Authors:

Dwayne A. Oslund

J. S. A. Clark

Approved by:

William E. Peterson

Thesis Advisor

Richard L. Elster

Second Reader

Richard L. Elster

Chairman, Department of Administrative Sciences

Kenneth T. Marshall

Dean of Information and Policy Sciences

ABSTRACT

The purpose of this study is to present analytic techniques for developing enlistment standards models which attempt to improve upon existing methods. Alternative criteria for measuring successful operational performance, and a means of measuring utility are also provided. Another purpose of this study is to discover if the Navy's system of selecting personnel for the Aviation Machinist's Mate (AM) rating may be improved.

Two criteria were utilized in developing these models--length of service, and a composite measure of success that considers length of service, highest paygrade achieved, and reenlistment eligibility. Measures on individual's at the time of enlistment are used as predictor and discriminating variables. Six models are developed and analyzed using regression and discriminant techniques. Utility analysis is conducted on each of these models as a means for measuring their usefulness in monetary terms. Recommendations for future research are also presented.

TABLE OF CONTENTS

I.	INTRODUCTION	9
	A. PURPOSE OF ANALYSIS	9
	B. RATING SELECTED FOR ANALYSIS	10
	C. ORGANIZATION OF THIS STUDY	11
II.	SELECTION OF VARIABLES	12
	A. SELECTION BACKGROUND	12
	B. REVIEW OF PREVIOUS MILITARY STUDIES	13
	C. CRITERION AND PREDICTOR VARIABLES	15
III.	DATA BASE DEVELOPMENT	16
	A. MASTER FILE	16
	B. AD DATA SET	17
	1. Screens	17
	2. Created Variables	19
IV.	STATISTICAL ANALYSIS	25
	A. DESCRIPTIVE ANALYSIS	25
	1. Frequency Analysis	25
	2. Summary Statistics	27
	3. Multivariate Correlation Analysis	28
	B. REGRESSION ANALYSIS	29
	1. Stepwise Regression	29
	2. Multiple Regression	32
	3. Validation	34
	C. DISCRIMINANT ANALYSIS	35
	1. Stepwise Discriminant Analysis	36
	2. Discriminant Analysis	36
V.	UTILITY ANALYSIS	43

A.	PURPOSE OF UTILITY ANALYSIS	43
E.	THEORY OF UTILITY ANALYSIS	44
1.	Model Validity	47
2.	Base Rate	47
3.	Selection Ratio	47
C.	ESTIMATING THE UTILITY OF A MODEL	48
D.	RESULTS OF UTILITY ANALYSIS	49
1.	Regression Models	50
2.	Discriminant Models	51
VI.	CONCLUSIONS AND RECOMMENDATIONS	54
A.	RESULTS	55
E.	RECOMMENDATIONS	57
APPENDIX A:	DATA BASE DEVELOPMENT PROGRAMS	59
APPENDIX B:	DESCRIPTIVE ANALYSIS RESULTS	75
APPENDIX C:	REGRESSION ANALYSIS PROGRAMS	84
A.	REQUIREMENTS AND ASSUMPTIONS	84
E.	STEPWISE REGRESSION	85
C.	LINEAR REGRESSION	85
APPENDIX D:	DISCRIMINANT ANALYSIS PROGRAMS	89
A.	REQUIREMENTS AND ASSUMPTIONS	89
E.	DISCRIMINANT ANALYSIS	90
APPENDIX E:	UTILITY ANALYSIS PROGRAMS	93
A.	CALCULATION OF CELL PROBABILITIES	93
1.	Regression Models	93
2.	Discriminant Models	95
E.	ESTIMATION OF CELL UTILITIES	96
C.	PROGRAMS USED TO CALCULATE UTILITIES	99
D.	CALCULATION OF BASE LINE UTILITIES	99
LIST OF REFERENCES	109

BIBLIOGRAPHY	111
INITIAL DISTRIBUTION LIST	112

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability (S)	
Dist	Special
A-1	



LIST OF TABLES

I.	Candidate Predictor/Discriminating Variables	26
II.	Selection Models	27
III.	Selected Summary Statistics	28
IV.	Predictors in Stepwise Regressions	31
V.	Regression Analysis Results	33
VI.	Regression Model Validities	34
VII.	Discriminant Analysis Results	40
VIII.	Relative Cell Utilities	49
IX.	Utility Results - Regression Models	52
X.	Utility Results - Discriminant Models	53
XI.	Program to Extract Data from the Master File	61
XII.	Program to Screen the AD Data	67
XIII.	Program to Create New Variables	69
XIV.	Selected Frequencies	76
XV.	Selected Correlations	81
XVI.	Sample Validation Program	87
XVII.	Sample Discriminant Analysis Program	92
XVIII.	Illustrative Actual and Predicted Scores	94
XIX.	Illustrative Discriminant Example	96
XX.	First Utility Analysis Program	101
XXI.	Partial Output from the First Utility Program	102
XXII.	Second Utility Analysis Program	103
XXIII.	Partial Output from the Second Utility Program	106
XXIV.	Third Utility Analysis Program	107

I. INTRODUCTION

For the remainder of this decade and beyond, the Navy is faced with the difficult problem of attracting and retaining sufficient personnel to meet its ever increasing manpower requirements. The fleet is expanding toward 600 ships while the available manpower pool of 17 to 21 year old men and women is projected to decline. Each year, millions of dollars are spent recruiting, training and maintaining enlisted personnel. Numerous enlistment standards models have been developed to improve the screening, selection and assignment processes for all Navy ratings. Continuing enlistment standards research is important since it may improve manpower planning, reduce attrition, enhance job performance, and lengthen careers. It is through such research that the ultimate goal of increased readiness at lower cost may be realized.

A. PURPOSE OF ANALYSIS

This study attempts firstly to improve upon the methodology presently utilized to develop enlistment standards models. In particular, different techniques for developing such models are presented, along with alternative criteria for measuring successful performance. A means of measuring the utility, or usefulness, of such efforts is also provided. An attempt has been made to present these methods in a clear manner so that those researchers who follow may more readily understand the process. The analysis expands upon the experience of numerous similar efforts, including several graduate theses prepared at the Naval Postgraduate School and many research projects conducted under the

auspices of the Navy Personnel Research and Development Center (NPRDC) and the Center for Naval Analyses (CNA).

The secondary purpose of this study is to discover if the selection standards for one particular Navy rating may be improved upon by analyzing data available at the time of enlistment. Most predictive models developed to date have focused on successful completion of technical training schools, or on attrition. This study is aligned with a more recent analytic trend of attempting to predict successful operational performance in the fleet. This approach is considered most appropriate in today's highly technical Navy. The tremendous cost, in terms of both dollars and time, associated with training and retaining Navy personnel demands maximum return on investment. By focusing on operational success to develop a larger, more experienced career force, there exists the potential to reduce the burden of recruiting and training new enlistees.

E. RATING SELECTED FOR ANALYSIS

To accomplish the above stated purposes, data pertaining to operational performance of personnel in the Aviation Machinist's Mate (AM) rating were analyzed. AMs are aircraft engine mechanics who inspect, adjust, test, repair and overhaul engines used in all Navy airplanes and helicopters. AMs also perform routine maintenance, prepare aircraft for flight, and assist in handling aircraft on the ground or aboard ships. They perform maintenance and servicing on either jet or reciprocating engines, and on subsystems such as fuel, oil, induction, compression, combustion, turbine and exhaust. Other AM functions include supervising maintenance, analyzing fuel and oil samples, keeping records, evaluating engine performance, and maintaining accessory components, drive systems and gear boxes.

Aviation Machinist's Mates are assigned primarily to Naval Aviation squadrons or to Aircraft Intermediate Maintenance Departments. These assignments may be either afloat or ashore. AIs may also be assigned as instructors in training activities, and they are eligible to volunteer for flight duty as aircrewmembers. [Ref. 1]

Presently, there are over 13,000 men and women assigned to the AD rating. Since ADs are assigned to virtually every Navy aviation unit, they represent a vital element in ensuring a high degree of aircraft readiness is maintained. As such, the overall mission effectiveness of Naval Aviation units is directly linked to the quality and performance of members of the AD rating.

C. ORGANIZATION OF THIS STUDY

This chapter has discussed the purpose of this study, and described the AD rating and its importance to the Navy. The next chapter will provide background information on enlistment standards research, and present in general terms the evolution of predictor and criterion variables that emerged during the development of the models contained in this research. Chapter III describes the data base and AD data set that provided specific measures of operational performance for analysis and model formulation. Chapter IV presents the three statistical analysis techniques employed in developing six enlistment standards models. Chapter V provides the method and results of the utility analysis conducted on these models. Utility analysis represents a means by which the usefulness of similar efforts may be measured. Chapter VI draws conclusions from the analysis and presents recommendations for further research.

II. SELECTION OF VARIABLES

This chapter gives a brief description of some of the selection procedures in use at the time of the data collection. The results of previous research on predicting job performance are reviewed and predictor and criterion variables that have been shown to be useful are identified.

A. SELECTION BACKGROUND

At the time the data used in this analysis was collected, the Navy considered a number of applicant characteristics to guide selection and classification decisions. These characteristics included education, high school degree status, age, number of dependents, scores on the 12 Armed Services Vocational Aptitude Battery (ASVAB) subtests, and some composite scores. The Armed Forces Qualification Test (AFQT) is one of these composite scores, and an applicant's score on the AFQT depended on the sum of his scores on the ASVAB subtests Arithmetic Reasoning (AR), Spatial Perception (SP), and Word Knowledge (WK). The AFQT score was reported as a percentile--a score of 80 meant that an applicant's total score on the three subtests was higher than the scores achieved by 79 percent of his peers. The AFQT percentile score was also used to classify the applicant into one of five mental categories or AFQT groups. For example, those with a score of 90 or better were classified in mental group I, and those with a score of 10 or less were classified as group Vs.

Another composite score is the Success Chances of Recruits Entering the Navy (SCREEN) score. This score is derived from the personal characteristics of age at entry,

years of schooling, whether or not the applicant had dependents, and AFQT percentile score. This composite score has been used by recruiters since October 1976 in determining an applicant's eligibility to enlist. [Ref. 2]

A final composite score that is used in classifying recruits to the AD rating is made up of the sum of the recruit's standardized scores on the ASVAB subtests Arithmetic Reasoning (AR), Electronic Information (EI), General Science (GS), and Mathematical Knowledge (MK). A minimum score of 190 on this composite was necessary for a recruit to enter the AD rating.

E. REVIEW OF PREVIOUS MILITARY STUDIES

Studies on predicting military job performance have mainly concentrated on the survivability of recruits through various stages of their military careers. These stages include recruit training, Class "A" School, first two years of enlistment and first term of enlistment.

Lurie used number of dependents, years of education and AFQT score to predict the performance of the Electrician's Technician (ETN) and Ship's Serviceman (SH) ratings. He found that for the SH rating, non-high school graduates with lower AFQT scores were promoted faster than those with higher scores, although AFQT score had no impact on first term survival. The AFQT score did aid in predicting advancement and survival for members of the ETN rating. [Ref. 3] In another study of eight year survival rates, Lurie found that education level was the most important predictor. Interestingly he also found that mental group I recruits had the worst record in surviving Class "A" School. [Ref. 4]

A study on the factors affecting first term survival and reentry behavior of Machinist's Mates (MM) and Boiler

Technicians (BT) was conducted by Fletcher in 1979. He found that BTs with greater than 11 years of education had greatly improved chances of surviving their first term of enlistment. For MMs, those in the lowest and highest mental groups had greater survival probability than others. For both ratings, older men had a higher probability of survival. In relation to reenlistment, those BTs with 12 or more years of education had a low probability of reenlistment, while with MMs, having a dependent increased the probability of reenlistment. [Ref. 5]

Studies of military job performance have also investigated the effect of the Delayed Entry Program (DEP) on survival. Lockman found that if recruit quality (as measured by SCREEN) and training guarantees were controlled for, those who were in the DEP for three or more months had the highest survival rates [Ref. 6]. Thomason found that DEP, age, education, recruit training location, race, number of dependents, mental group and follow on tour assignments had varying degrees of significance in determining first term survivability [Ref. 7].

More recent studies have favored the use of multiple, rather than single measures of job performance. This is because it is rare that a single measure adequately covers the full scope of job performance. One approach has been to expand the survivability criteria to include other measures of job performance, such as eligibility to reenlist and the achievement of certain paygrades. A continuous criterion is not available under this approach; sailors are either categorized as a success or as a failure. Nesbitt [Ref. 8] and Snyder and Bergazzi [Ref. 9] defined various degrees of success or failure in their studies in an effort to generate more variability on the criterion.

C. CRITERION AND PREDICTOR VARIABLES

In most cases when a single job performance measure (criterion) has been used in previous research, a measure of survival has been the overwhelming choice. This is because such a criterion is readily measurable, is continuous, and is of importance to the Navy since the costs associated with attrition and subsequent replacement are considerable. Other single criteria have been length of service (IOS), time to promotion, highest rank or grade achieved, retention (as measured by reenlistment choice), and performance at Class "A" Schools.

The common predictors of job performance are education, number of dependents, age, sex, race, ASVAB subtest scores, AFQT scores, mental group, DEP status, and some "after accession" variables such as recruit training location, subsequent dependent status, and initial and follow on duty assignments.

In this study two criteria will be considered. The first will be an IOS criterion and the second will be a composite criterion where success will be defined as completing the first term of enlistment, being eligible for reenlistment, and achieving the paygrade E-4. The candidate predictor variables will be age at entry, sex, race, entry paygrade, education, dependent status, term of enlistment, ASVAB subtest scores, AFQT scores and the composite score to qualify for the AD rating. The specific variables from the AD data set used for analysis, as well as the evolution of the data set, are discussed in the next chapter.

III. DATA BASE DEVELOPMENT

This chapter provides information concerning the master data base and the subset of this master file, the AD data set, that was used in this study. The generation of this AD data set is described in detail, as are the specific predictor and criterion variables discussed in the previous chapter.

A. MASTER FILE

Enlisted history records on over 206,000 non-prior service sailors who entered the Navy during the period 1 September 1976 to 31 December 1978 were compiled by the Defense Manpower Data Center (DMDC) staff. This master file was created by merging data elements from four separate files: (1) the DMDC Cohort file, which is itself a compilation of information from DMDC's Enlisted Master Record and Loss files; (2) a Navy Health Research Center (NHRC) file; (3) a promotion examination file; and (4) a Chief of Naval Education and Training (CNET) file.

The DMDC Cohort file contains personal and demographic data on individuals at the time they entered the service. Additionally, it is updated quarterly by the Military Personnel Commands with active duty or service separation information as appropriate. This file provided over 150 variables to the master file.

The NHRC file contains information on each enlisted member of the Navy who has been or still is on active duty. It is updated quarterly from Navy Military Personnel Command (NMPC) change tape extracts, and follows a service member from date of enlistment to date of discharge. The NHRC file represents approximately 30 variables in the master file.

The promotion examination file includes advancement exam and promotion data, and the CNET file contains information on formal training received by individuals in the data base. Together these files provided over 60 variables to the master data base.

The master file, containing 243 variables, is maintained at the Naval Postgraduate School. The final update to the file includes DMDC data current as of 30 September 1982, and NHRC data current as of July 1982. The program to access data on the selected rating is contained in Appendix A.

E. AD DATA SET

This section describes the evolution of the AD data set that contains the observations and measures analyzed in this study. The AD data set was derived through a number of iterative screens, and new variables were created, in order to remove aberrant observations and to refine the predictor and criterion variables prior to statistical analysis. These iterative steps ultimately reduced the number of cases in the AD data set from 5,562 to 2,820 observations. The programs used to screen the data and to create new predictor and criterion variables are contained in Appendix A. The logic for these processes is discussed in the remainder of this chapter.

1. Screens

Since one purpose of this study was to analyze Aviation Machinist's Mates who had operational experience in the fleet, the first screen performed on the data was to select only those cases whose final DMDC rate (DMICRATE)¹ appeared in the last master file update as ADs. This means

¹The actual variable name associated with the comment is provided in parentheses throughout the remainder of this chapter.

they were working in the AD rating at either the time of their separation from the service, or at the time of the last file update if they were still on active duty. This screen allowed for people to migrate into the AD rating while ensuring that those cases who left for another rating were excluded from the analysis.

The cases were next screened for ADS with no prior Navy service (PRIORSEV). In addition, individuals who may have changed their social security number (SSNCHNGE) were removed from the sample. These screens ensured that no observations were counted more than once in the analysis.

The observations in the AD data were then screened to select only those people who were tested on ASVAE Forms 5, 6, or 7 (TESTFORM) at enlistment. These test forms were in use during the period in which the individuals in the data set enlisted. Also, those cases whose subtest scores (ASVAEs) were impossibly high were eliminated from the data set.

Two other screens were conducted to capture those observations who enlisted in either the Regular Navy or Naval Reserve (SERVACCS), and who were known to have signed up for at least four years active service (RECENLST). It is worth noting that during model development, the term of enlistment measure (TERMENLT) was consistently significant, but with a negative value for the parameter estimate. This indicated that those individuals who enlisted for longer obligated service actually served less time than those who signed up for shorter terms of enlistment. The parameter estimate for term of enlistment changed to a positive value when the RECENLST screen was implemented. Apparently, by screening for four year active duty obligors, the data set then excluded those reservists who were required to serve only three years of their six year obligation on active duty. For these observations, a six year term of enlistment

was an erroneous value for the TERMENLT variable. This important discovery reveals a probable flaw in earlier similar enlistment standards analytic efforts.

Another screen facilitated inclusion of those cases for which there was clear indication of their eligibility to reenlist (ELGREUP1 or ELGREUP3). The final screen in setting up the AD data set included only those separated individuals who could be easily identified by "good" or "bad" interservice separation codes (ISC). Observations with separation codes in the "grey" area (death, hardship discharge, entry into officer programs, or medical disqualification) were removed from the data set since it was felt a legitimate determination of their success or failure could not be made.

Having incorporated these screens, frequency distribution analysis facilitated the removal of aberrant or impossible cases. For example, the maximum length of service between 1 September 1976 and 30 September 1982, the period of the data base, was 72 months. Cases who were listed as having greater than 72 months LOS were removed from the data set.

2. Created Variables

This discussion identifies the variables created in addition to those already in the master data base. Creating these variables facilitated more detailed analysis of observations in the AD data set, and enhanced the enlistment standards model development process. The following comments will also address the reasons for selecting some variables and not others.

a. Predictor Variables

There were several ways that individuals in the master data base might appear in the AD rating during their

career. They may have enlisted in a program to become an AD, taken the AD rating exam, and/or achieved the AD rating through on the job training. To distinguish between the various combinations of these processes, an entry group variable (ENTRYGRP) was created. Frequency analysis of this new variable confirmed that the final DMDC rate of AD screened and selected only those cases who actually ended up as ADs. An effort was made to develop models for various combinations of these entry groups during stepwise regression analysis. However, the derived models were not significant, and they did not improve upon the models ultimately selected for analysis.

A common predictor variable in enlistment standards models is one dealing with education. The measure in the master data base reflecting education level (HYEC) was converted from a qualitative value to a continuous variable (CHYEC) to facilitate statistical analysis. In addition, a dichotomous (0,1) variable was created to reflect attainment of a high school degree (HSDG). During stepwise analysis, which is discussed in the next chapter, each of these two new variables was offered separately as a candidate predictor variable. In nearly every instance, HSDG was shown to be more significant than CHYEC.

Other common predictor variables which measure entry-level attributes are ASVAB subtest scores. To allow the use of these measures of individual characteristics, the scores were standardized, and the created variables (SASVABs) were considered during model development. As mentioned in Chapter II, standardized ASVAB subtest scores are used in various combinations as composite measures. One of these composite variables is AFQT percentile (AFQTFCNT), which also yields AFQT groups (AFQTGRPS). Another composite is the score used to determine eligibility for the AD rating. Two variables were created in the AD data set to

identify this latter composite measure. The first variable created (ADCOMPOS) was a continuous variable which had a value equal to the sum of the four ASVAB standardized subtest scores that make up the composite. The second variable created (ADMINSCR) was a dichotomous variable which distinguished those ADCOMPOS values greater than or equal to 190 from those ADCOMPOS values less than 190. Each time one of these four composite measures was offered as a candidate predictor variable during regression model development, three separate trials were run. One trial contained the composite measure and all 12 SASVAB variables. Another trial contained the composite variable and only those SASVAB variables that did not make up the composite variable. The third trial contained only the composite measure with no SASVAB variables. Additionally, the trials contained either AFQTFCNT or AFQTGRPS, and either ADCOMPOS or ADMINSCR. The purpose of this iterative process was to ensure multicollinearity effects were minimized among the independent variables. During the development of the regression models, AFQTFCNT and ADMINSCR were consistently shown to be more significant than AFQTGRPS and ADCOMPOS respectively. For this reason, they were included among the final candidate predictor variables used in stepwise regression analysis.

Another predictor variable commonly considered by enlistment standards research deals with marital status and dependents. The master file contains a qualitative variable (MRTLDPND) which reflects marital status and number of dependents. This study created a dichotomous variable (DEPENDIS) which distinguishes single individuals from those who are married and/or who have children. Again an iterative process revealed this created variable to consistently be more significant.

The effects of race and sex were also considered in the analysis by creating new variables. The best

variable in the master file to indicate race and ethnic status identified categories of whites, blacks and others (RACE). Since this variable was qualitative, three dummy variables were created (WHITE, BLACK, and OTHER). To allow analysis of the effects of sex, the master file variable (SEX) was converted to a "0,1" variable (NUSEX).

Several other predictor variables were considered and tested for significance and possible inclusion in the final set of candidate predictor variables prior to developing the regression models. Age at enlistment (ENTRYAGE), enlistment paygrade (ENTRPAYG) and term of enlistment (TERMENLT) were among those selected. Many variables were rejected because other measures were better able to capture the desired effects. One particular variable which did not show to be significant was the composite SCREEN variable (SCREEN) discussed in Chapter II. This may be because the components of the SCREEN variable are individually more appropriate for analysis, particularly when the emphasis is on predicting operational performance in the fleet. Similar results were cited by McGarvey [Ref. 10].

The final set of predictor variables created in the AD data set are interaction terms. These variables represent all two-level interactions of the seven variables that met the specified significance level during stepwise regression analysis. The development of these variables is discussed in more detail in Chapter IV.

b. Criterion Variables

As discussed in Chapter II, this study used two criterion variables when developing the six models--length of service measures and success measures. The length of service measure for regression models is a continuous variable (TAFMS1), and for discriminant models is a dichotomous

variable (SUCCTAF).² SUCCTAF was assigned a value of 'one' if the value of TAFMS1 was greater than or equal to 48 months, or if the value of TAFMS1 was greater than or equal to 45 months and the individual entered the Navy in October, November or December 1978 (LATEENLT). This was done to ensure those cases who did not have the opportunity to serve 48 months were not improperly classified as failures.

Individuals were considered as successes, for purposes of this analysis, if they served 48 months or longer, achieved paygrade E-4, and were recommended for reenlistment. Again, observations who did not have the opportunity to serve 48 months were also considered successful on the ICS portion of this criterion if they served at least 45 months. The success criterion variable (SUCCESS2) captures these measures by considering SUCCTAF and two other created variables (SUCCPAYG and SUCCREUP).

SUCCPAYG identifies those cases who achieved E-4 as measured by two created variables (PAYGRADE and NUHPAY). PAYGRADE was created from one of two DMDC variables (PAYGRDE1 or PAYGRDE3) that measure an individual's paygrade at the last file update or upon discharge from the service, as appropriate. NUHPAY was created by converting an NHRC variable (HPAYGRD) from a categorical to a numeric variable. Using both DMDC and NHRC measures of paygrade ensured correct classification of an individual on this portion of the criterion.

SUCCREUP, the eligibility to reenlist portion of the success criterion, was derived from the DMDC variable (ELGREUP1) that captured the reenlistment code assigned upon an individual's discharge from the service. Service members on active duty as of the last master file update were considered eligible to reenlist, as long as there was no

²Discriminant analysis requires the use of categorical vice continuous variables as classification variables.

other information to the contrary. The next chapter will discuss how the information contained in the AD data set was specifically evaluated using three separate statistical analysis techniques.

IV. STATISTICAL ANALYSIS

Three distinct statistical methods were employed in this research: Descriptive Analysis, Regression Analysis and Discriminant Analysis. All methods used Statistical Analysis System (SAS) procedures to analyze the data and develop the models. Table I contains a list of the 46 candidate predictor/discriminating variables used in this study. In all, six sets of variables emerged, and each set was analyzed using both regression and discriminant techniques for comparison. These six sets of predictor/discriminating variables are shown in Table II, along with the appropriate criterion/classification variable. Each method, along with the results, are discussed in the following sections of this chapter. It is worth noting that, while the results may not represent a marked improvement over the selection process in use when individuals in the data set enlisted, the methodology presented may be applied to further analysis of the AD rating or to any other rating in the Navy.

A. DESCRIPTIVE ANALYSIS

Descriptive analysis was accomplished through review of frequency distributions, summary statistics and multivariate correlations.

1. Frequency Analysis

Frequency distributions are summary tables in which data are grouped or arranged into conveniently established numerically ordered classes or categories. The process of data analysis is, therefore, made much more manageable and

TABLE I
Candidate Predictor/Discriminating Variables

Variable	Label
AFCTECNT	AFCT PERCENTILE (CR EQUIVALENT)
AFCTGFEFS	AFCT GROUPS (5, 4C, 4B, 4A, 3B, 3A, 2, 1)
ENTFYAGE	AGE OF INDIVIDUAL AT TIME OF ENTRY
ENTFFAIG	ENTRY PAYGRADE (EO--011)
TERMENIT	TERM OF ENLISTMENT (NO. OF YEARS)
ESDCG	HIGH-SCHOOL GRADUATE (1) V. OTHER (0)
DEPENCLS	SINGLE, NO DEPENDENTS (0), OTHERWISE (1)
CHVEE	CCNVEED NUMBER OF YEARS OF EDUCATION
SASVAEGI	STANDARDIZED SCCRE - GENERAL INFORMATION
SASVAENO	STANDARDIZED SCCRE - NUMERICAL OPERATIONS
SASVAEAD	STANDARDIZED SCCRE - ATTENTION TO DETAIL
SASVAEWK	STANDARDIZED SCCRE - WORD KNOWLEDGE
SASVAEAR	STANDARDIZED SCCRE - ARITHMETIC REASONING
SASVAESP	STANDARDIZED SCCRE - SPACE PERCEPTION
SASVAEMK	STANDARDIZED SCCRE - MATH KNOWLEDGE
SASVAEEI	STANDARDIZED SCCRE - ELECTRONIC INFO
SASVABMC	STANDARDIZED SCCRE - MECH COMPREHENSION
SASVAEGS	STANDARDIZED SCCRE - GENERAL SCIENCE
SASVAESI	STANDARDIZED SCCRE - SHOP INFORMATION
SASVAEAI	STANDARDIZED SCCRE - AUTO INFORMATION
ELACK	(1) BLACK, ELSE (0)
CTEER	(1) NEITHER BLACK NOR WHITE, ELSE (0)
NUSEX	(1) MALE, (0) FEMALE
ACCCMECS	AC ASVAB COMPOSITE
ADMINSCR	AL ASVAB COMPOSITE SCREEN
INTER01	DEPENDENTS * HSDG
INTER02	DEPENDENTS * BLACK
INTER03	DEPENDENTS * NUSEX
INTER04	DEPENDENTS * TERMENIT
INTER05	DEPENDENTS * SASVABAI
INTER06	DEPENDENTS * ADMINSCR
INTER07	HSDG * BLACK
INTER08	HSDG * NUSEX
INTER09	HSDG * TERMENIT
INTER10	HSDG * SASVAEAI
INTER11	HSDG * ADMINSCR
INTER12	BLACK * NUSEX
INTER13	BLACK * TERMENIT
INTER14	BLACK * SASVAEAI
INTER15	BLACK * ADMINSCR
INTER16	NUSEX * TERMENIT
INTER17	NUSEX * SASVAEAI
INTER18	NUSEX * ADMINSCR
INTER19	TERMENIT * SASVABAI
INTER20	TERMENIT * ADMINSCR
INTER21	SASVAEAI * ADMINSCR

meaningful. In this study, frequency analysis was performed to provide counts and percentage distributions of individuals in the sample, and to illustrate the range of the predictor and criterion variables. This information provided a base upon which to screen aberrant observations and to compare the results of this study. Frequency distributions are provided in Appendix B for the AD rating.

TABLE II
Selection Models

Model	Predictors/ Discriminating Variables	Regression Criterion Variable	Discriminant Classification Variable
A	DEPENDTS TERMENLT ADMINSCR HSDG ELACK CHER NUSEX	TAFMS 1	SUCCTAF
E	TERMENLT INTER03 INTER04 INTER08 INTER14 INTER21	TAFMS 1	SUCCTAF
C	INTER03 INTER08 SASVAEWK ENTERPAYG	TAFMS 1	SUCCTAF
D	DEPENDTS HSDG CTHER TERMENLT SASVAEAI SASVABWK	SUCCESS2	SUCCESS2
F	INTER03 INTER09	SUCCESS2	SUCCESS2
F	INTER03 INTER09 INTER21 CHER SASVAEEI SASVABMK SASVAESI AFCTGRPS CHYEC	SUCCESS2	SUCCESS2

Note: Variable sets A, E, D and E resulted from stepwise regression techniques. The variable sets C and F resulted from stepwise discriminant techniques. Table I provides the labels for these variables.

2. Summary Statistics

Like frequency distributions, descriptive summary statistics are useful for analyzing and interpreting quantitative data. These summary statistics represent properties of location, dispersion and shape, and may be used to extract and summarize features of the data set. Representative summary statistics for variables in the AD data set are shown in Table III.

TABLE III
Selected Summary Statistics

VARIABLE	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE
AFCTECNT	43.45	20.50	6.00	99.00
TAFMS1	49.22	9.44	2.00	72.00
ENTRYAGE	18.85	1.62	17.00	30.00
SASVAEGI	49.91	7.71	20.00	66.00
SASVAENC	50.43	7.80	23.00	69.00
SASVAEAD	50.34	9.28	20.00	80.00
SASVAEWK	48.20	7.51	30.00	64.00
SASVAEAF	48.92	6.98	29.00	65.00
SASVAESP	49.12	8.39	21.00	66.00
SASVAEMK	50.46	7.01	26.00	67.00
SASVAEEI	51.04	6.58	20.00	68.00
SASVAEMC	50.08	8.24	25.00	71.00
SASVAEGS	49.57	7.14	24.00	70.00
SASVAESI	50.90	8.48	20.00	65.00
SASVAFAI	51.16	9.51	26.00	67.00
CHYEC	11.79	0.73	3.50	16.00
ALCCMFCS	159.95	19.19	99.00	264.00

3. Multivariate Correlation Analysis

Measuring the strength of the relationship between variables may be accomplished by correlation analysis. This technique enables one to gain an idea of the degree of association or covariation that a variable has with another variable. The summary measure that expresses the extent of this relationship is the coefficient of correlation, r , whose values range from -1 for perfect negative correlation to +1 for perfect positive correlation. Values close to zero indicate little systematic covariation between two variables. Correlation coefficients for quantitative variables used in this study are contained in Appendix B.

Assessing the strength of association between variables does not allow a researcher to predict the value of one variable from the value of another variable. The latter involves regression techniques, and is presented in the next section of this study.

E. REGRESSION ANALYSIS

Regression analysis is one method used to develop a statistical model that can predict the values of a dependent or response variable based on the values of independent or explanatory variables. Rather than merely measuring the association between variables with correlation analysis, a regression model attempts to predict or explain the value of the criterion variable by developing an equation that is based on weighted values of one or more predictor variables.

In developing the selection models in this study, the process employed was to first apply a variable "search" procedure called stepwise regression. The resultant models were then analyzed by simple regression analysis, and validated against a hold-out sample of the data set. The details of this process, the specific models derived, and results of the analysis are reported in the following sections. Appendix C contains a discussion of regression analysis assumptions and methodology.

1. Stepwise Regression

One of the desired characteristics of a regression model is parsimony, which means including the least number of explanatory variables that permit adequate interpretation of the dependent variable of interest. Such models are easier to interpret and are not as likely to be affected by multicollinearity³ problems. In developing the models for this study, stepwise regression procedures were employed to find a "best" combination of predictor variables, thereby avoiding the computationally complex and costly process of examining all possible regressions.

³Multicollinearity refers to the condition in which some of the independent variables are highly correlated with each other. When multicollinearity is present, the values of the regression coefficients for the correlated variables may fluctuate dramatically.

In this study, two sets of candidate predictor variables were analyzed with the stepwise procedure. The first set included those entry-level attributes and measures that were considered likely to be good predictors of each criterion, based on a review of similar enlistment standards studies. As discussed in Chapter II, these variables included individual and demographic measures such as mental ability, amount of education, entry age, entry paygrade, marital status, AFQT percentile, and ASVAB scores. Table IV provides a list of the 18 candidate variables from the AD data set that were used in the stepwise procedure.

The second set of candidate predictor variables included the seven variables from the first set that met the specified significance level for inclusion in the stepwise model. In addition, this set included all two-level interactions* of these seven variables. Inclusion of interaction terms in this study represents a marked departure from previous enlistment standards research. The results of this analysis clearly indicate the presence of interaction effects among predictor variables. The seven predictor variables and 21 interaction terms used in the stepwise analysis are also contained in Table IV.

Using these two sets of candidate predictor variables, the stepwise procedure was run on each of the two criterion variables, TAFMS1 and SUCCESS2, which were defined in Chapter III. The resulting four models were developed from a uniform random split, the derivation sample, of 1440 observations in the AD data set. This derivation sample constituted approximately half of the 2820 total cases in the AD data set. So doing facilitated cross-validation of

*An interaction involves the product of two or more independent variables, and is included in a regression model when the relationship between one independent variable and the dependent variable changes for differing values of another independent variable [Ref. 11].

TABLE IV
Predictors in Stepwise Regressions

Variable	Label
-- FIRST SET --	
AFCTPCNT	- AFQT PERCENTILE (CR EQUIVALENT)
ENTFYAGE	- AGE OF INDIVIDUAL AT TIME OF ENTRY
ENTFPAYG	- ENTRY PAYGRADE (EO--011)
TERMENIT	- TERM OF ENLISTMENT (NO. OF YEARS)
ESLG	- HIGH-SCHOOL GRADUATE (1) V. OTHER (0)
DEFFENDIS	- SINGLE, NO DEPENDENTS (0), OTHERWISE (1)
SASVAEGI	- STANARDIZED SCORE - GENERAL INFORMATION
SASVAENO	- STANARDIZED SCORE - NUMERICAL OPERATIONS
SASVAEAD	- STANARDIZED SCORE - ATTENTION TO DETAIL
SASVAEWK	- STANARDIZED SCORE - WORD KNOWLEDGE
SASVAESP	- STANARDIZED SCORE - SPACE PERCEPTION
SASVAEMC	- STANARDIZED SCORE - MECH COMPREHENSION
SASVAESI	- STANARDIZED SCORE - SHCP INFORMATION
SASVAEAI	- STANARDIZED SCORE - AUTO INFORMATION
ELACK	- (1) BLACK, ELSE (0)
CTEEB	- (1) NEITHER BLACK NOR WHITE, ELSE (0)
NUSEX	- (1) MALE, (0) FEMALE
ADMINSCR	- AD ASVAB COMPOSITE SCREEN
-- SECOND SET --	
TERMENIT	- TERM OF ENLISTMENT (NO. OF YEARS)
ESLG	- HIGH-SCHOOL GRADUATE (1) V. OTHER (0)
DEFFENDIS	- SINGLE, NO DEPENDENTS (0), OTHERWISE (1)
SASVAEAI	- STANARDIZED SCORE - AUTO INFORMATION
ELACK	- (1) BLACK, ELSE (0)
NUSEX	- (1) MALE, (0) FEMALE
ADMINSCR	- AD ASVAB COMPOSITE SCREEN
INTER01	- DEPENDITS * HSDG
INTER02	- DEPENDITS * ELACK
INTER03	- DEPENDITS * NUSEX
INTER04	- DEPENDITS * TERMENIT
INTER05	- DEPENDITS * SASVAEAI
INTER06	- DEPENDITS * ADMINSCR
INTER07	- HSDG * BLACK
INTER08	- HSDG * NUSEX
INTER09	- HSDG * TERMENIT
INTER10	- HSDG * SASVAEAI
INTER11	- HSDG * ADMINSCR
INTER12	- ELACK * NUSEX
INTER13	- ELACK * TERMENIT
INTER14	- ELACK * SASVAEAI
INTER15	- ELACK * ADMINSCR
INTER16	- NUSEX * TERMENIT
INTER17	- NUSEX * SASVAEAI
INTER18	- NUSEX * ADMINSCR
INTER19	- TERMENIT * SASVAEAI
INTER20	- TERMENIT * ADMINSCR
INTER21	- SASVAEAI * ADMINSCR

the models against a hold-out sample, the validation sample, whose characteristics would not influence the original

development of the models. The predictor variables that remained in the model at the termination of the stepwise procedure were significant at $p < .10$, and most variables were significant at $p < .05$. The four models themselves were significant at $p < .0001$.

2. Multiple Regression

The four models developed by the stepwise process were next analyzed using the SAS Regression procedure to describe the particular straight line model that best fit the data. Table V contains the printed output from the SAS Regression procedure that was run on each of the four models. For comparative purposes, two models developed by discriminant analysis techniques, discussed in the next section of this chapter, are also shown in table V. The SAS User's Guide provides a detailed description of the statistics that are included in the tables, as well as their method of computation [Ref. 12]. It can be seen that Model E, with the highest R-SQUARE and all variables statistically significant, is the preferred regression model.

The proportion of variation in the criterion variable explained by the set of predictor variables selected is called the coefficient of multiple determination, and is denoted R-SQUARE. The values of R-SQUARE for the models developed in this study are relatively low. This may be partially attributable to the large number of observations in the AE data set. However, it is also likely that the variation of the criterion variable, length of service or success as defined in this study, is also due to factors not captured by the entry-level attributes and measures used as predictor variables. These factors, which affect an individual's performance and decision to remain in the service, present themselves subsequent to enlistment. They may include satisfaction with initial assignment, geographical

TABLE V
Regression Analysis Results

Model	Predictors	Parameter Estimates	Prob > T	R Square	F Value
A	INTERCEPT	29.049	0.0001	0.0537	11.613
	DEFENDTS	2.841	0.0636		
	TERMENLT	3.639	0.0001		
	ADMINSCR	-1.207	0.0260		
	ESDGG	1.807	0.0036		
	CTEER	2.254	0.0294		
	NUSEX	4.171	0.0079		
	ELACK	1.729	0.0131		
B	INTERCEPT	32.140	0.0001	0.0547	13.828
	TERMENLT	3.890	0.0001		
	INTER03	15.724	0.0026		
	INTER04	-2.937	0.0173		
	INTER08	2.113	0.0004		
	INTER14	0.032	0.0398		
	INTER21	-0.024	0.0134		
C	INTERCEPT	31.746	0.0001	0.0220	8.089
	INTER03	3.888	0.0163		
	INTER08	2.137	0.0004		
	SASVABWK	-0.101	0.0022		
	ENTFPAYG	0.416	0.3685		
D	INTERCEPT	0.535	0.0002	0.0255	6.238
	DEFENDTS	0.172	0.0131		
	TERMENLT	0.053	0.0549		
	ESDGG	0.115	0.0001		
	CTEER	0.080	0.0871		
	SASVABAI	0.001	0.5630		
	SASVABWK	-0.003	0.1028		
E	INTERCEPT	0.663	0.0001	0.0198	14.501
	INTER03	0.196	0.0064		
	INTER09	0.030	0.0001		
F	INTERCEPT	0.565	0.0309	0.0370	6.107
	INTER03	0.202	0.0053		
	INTER09	0.038	0.0001		
	INTER21	-0.001	0.0576		
	CTEER	0.101	0.0297		
	SASVABEI	0.006	0.0022		
	SASVABSI	0.002	0.1456		
	CHYEC	-0.033	0.1138		
	AFQTGRPS	-0.027	0.0092		

location of duty assignment, command climate, unit employment, change in marital status, societal values and pressures, and educational and economic opportunities outside the military. These factors or measures are post hoc considerations that are not available when screening candidates for enlistment and initial rating assignment. They

are issues that are appropriate for more sophisticated methodologies, for example, covariance structure analysis which can treat complicated enlistment standards models as a series of simultaneous equations that capture performance as a "multiple-stage" process occurring throughout an individual's military career. [Ref. 10]

3. Validation

The results of the regression procedure were next validated against the hold-cut sample. Each of the regression models was derived from a uniform random sample, the derivation sample, of the observations in the AD data set. The SAS Regression procedure was employed to calculate the parameter estimates for the associated predictor variables using data from observations in this derivation sample. The SAS Score procedure then used these estimates to predict the value of the criterion variable for each observation in the validation sample. Finally, these predicted values were correlated with the actual values of the criterion in the validation sample. These correlations represent the validation coefficients for each model, and are shown in Table VI.

TABLE VI
Regression Model Validities

Model	First Validity Coefficient	Second Validity Coefficient	Average Validity
A	0.21342	0.20317	0.21
E	0.21536	0.21683	0.22
C	0.14455	0.13612	0.14
L	0.17387	0.13766	0.16
F	0.17790	0.12751	0.14
F	0.14430	0.13531	0.14

Note: The First Validity Coefficient is the result of the cross-validation, and the Second Validity Coefficient results from the double cross-validation. The reported average is the simple arithmetic mean.

As a further check of the validity of the six regression models, the process was repeated by deriving parameter estimates from the validation sample, and using these estimates to correlate the actual and predicted values of the criterion for observations in the derivation sample. This double cross-validation technique is described in detail by Campbell [Ref. 13]. Table VI also contains this second set of validity coefficients for the six models.

Occasionally, concern is expressed that random samples may not be from a homogeneous population, and, therefore, the sample correlations may differ from the population correlations. One method of addressing the problem of heterogeneous samples is to average the correlation coefficients to obtain a single estimate of the population correlation. If the sample correlations are of about the same value and if they are not too large, as is the case with this study, this simple arithmetic mean will suffice. Were this not the case, however, another technique is to use transformations to Fisher's z coefficients. [Ref. 14] The simple arithmetic average correlations are also presented in Table VI. Appendix C contains the program used to calculate validity coefficients.

C. DISCRIMINANT ANALYSIS

The third statistical method employed in this research was discriminant analysis. The use of discriminant analysis allows observations to be classified into two or more groups or categories on the basis of one or more numeric variables. As was done with regression analysis, the discriminant models were derived and analyzed from the derivation sample of the data set, and tested against the hold-out sample of observations. Variables in the model were again selected using stepwise techniques. The resulting two models, and

the four models developed by regression analysis, were then analyzed using the SAS Discriminant procedure. The program used in this analysis is contained in Appendix D, along with a discussion of discriminant analysis assumptions and methodology.

1. Stepwise Discriminant Analysis

The SAS Stepwise Discriminant procedure was employed to select the most useful discriminating variables. It is a logical and efficient method of choosing an optimal combination of variables. Their selection to enter or leave the model is based on either the significance level of an F test or a squared partial correlation criterion. The selected variables are those which contribute most to the discriminatory power of the model. [Ref. 12]

The variables chosen by the stepwise discriminant process were selected from the 46 candidate variables shown previously in Table I. The entry-level attributes and measures that were considered likely to be good predictors, as discussed in Chapter II, represent 25 of these candidate variables. The other 21 variables are the two-level interaction terms considered during regression analysis of the AD data set. The procedure was run on each of the two criterion variables, SUCCIAF and SUCCESS2, discussed in Chapter III. The criterion variables define the groups into which each observation will be classified by discriminant analysis, and are called classification variables.

2. Discriminant Analysis

As previously mentioned, discriminant analysis involves the study of differences between two or more groups, defined by a single nominal level variable, with a set of common discriminating variables.

The SAS Discriminant Analysis procedure provided the means for conducting discriminant analysis of the AD data set. The procedure was run on each of the six models developed by stepwise regression and stepwise discriminant processes. Each observation is placed in the class from which it has the smallest generalized squared distance. Also taken into account were the prior probabilities of group membership. These probabilities are obtained from a frequency distribution of actual successes and failures of the sample data set. This was considered appropriate since this study is attempting to improve upon the selection process in use at the time the individuals enlisted.

Table VII contains the results of discriminant analysis. Each procedure incorporated the prior probability of group membership, indicated on the classification matrix as PRIORS. The classification matrix is divided into four elements which depict the number of actual (row) versus predicted (column) classifications into successful (1) or failure (0) groups. The four elements (actual, predicted) in the matrix are:

- (0,0) The number of failure cases predicted to be failures
- (1,0) The number of successful cases predicted to be failures
- (0,1) The number of failure cases predicted to be successful
- (1,1) The number of successful cases predicted to be successful

Each section first contains the classification matrix developed by applying the classification function to the derivation sample. The second classification matrix depicts the results of applying this same classification function to observations in the hold-out sample, thereby validating the model.

The table also shows two rates relevant to each classification matrix. The first rate is the percentage of correct classifications, called the "hit rate", which provides a measure of the accuracy of the discriminant model. The second rate is the percentage of enlistees who were classified as (1,1) compared to all cases who were predicted as successful. It is called the "success rate", and it provides a measure of how well this selection model would have performed. It may be compared to the original selection strategy success rate, the priors. Success rate is an important consideration with utility analysis, and will be addressed further in Chapter V. As with regression analysis, Model B is again the preferred model since it is the only one that improves upon the selection strategy in existence during the timeframe of the AD data set.

To illustrate how the results may be interpreted, an example of the classification matrices for Model A will be explained. The model correctly classified 49 observations as failures and 1079 observations as successful. The sum of these correct classifications represents 79 percent of the total of 1440 observations in the derivation sample. To test the model's accuracy, the classification function is applied to the validation sample. The second classification matrix indicates 47 failure and 1039 successful observations were correctly classified. The sum represents a hit-rate of 79 percent of the total of 1380 observations in the hold-cut sample. The consistency of these hit-rates indicates the model is valid. The model betters the 85 percent success rate experienced by the Navy with the selection process used at the time the observations enlisted.

However, it is difficult to significantly improve upon such a high success rate. Additional entry-level attributes and measures might be found to better capture success as defined in this study. An alternate approach

would be to redefine the success criterion. In either case, however, the methodology presented in this chapter may be similarly followed to develop and test enlistment standards models. The next chapter will discuss a method by which the utility of such an effort may be measured.

TABLE VII
Discriminant Analysis Results

Model	Errors C	1	Classification Matrix				Hit Rate	Success Rate	
A	0.15	0.85	Predicted SUCCIAF				0.78	0.87	
			C	1	Total				
			Actual SUCCIAF	0	49	161	210		
				1	151	1079	1230		
			Total	200	1240	1440			
			Predicted SUCCIAF				0.79		0.88
			C	1	Total				
			Actual SUCCIAF	0	47	136	183		
				1	158	1039	1197		
			Total	205	1175	1380			
B	0.15	0.85	Predicted SUCCIAF				0.85	0.85	
			C	1	Total				
			Actual SUCCIAF	0	1	209	210		
				1	2	1228	1230		
			Total	3	1437	1440			
			Predicted SUCCIAF				0.87		0.87
			C	1	Total				
			Actual SUCCIAF	0	0	183	183		
				1	1	1196	1197		
			Total	1	1379	1380			

V. UTILITY ANALYSIS

This chapter contains an explanation of the applicability of utility analysis to the development of selection procedures, and discusses the theory of utility analysis. The methodology used in this study to apply utility analysis is described, along with sections on the calculation of cell probabilities for regression and discriminant models, and a section on estimating cell utilities. More detail on the calculations and programs used for utility analysis may be found in Appendix E.

A. PURPOSE OF UTILITY ANALYSIS

The development of a model for use in predicting an applicant's future performance in a particular job is a very necessary part of most selection procedures. However, the model itself does not constitute enough information to enable a decision to be made on whether or not it is worth implementing. The validity of the model is one indicator of its potential usefulness but, as will be seen, other factors significantly affect the usefulness of a model. All organizations would find it valuable to be able to judge the worth of their strategy in quantitative terms, particularly when comparing their existing strategy to a newly developed, competing strategy. A framework is needed which will allow the evaluation of a selection model in terms of the institutional gains (or losses) that are expected to result when that model is used to guide decisions on selection. Classical utility analysis provides such a framework, and it allows the calculation of usefulness to be made in terms of actual dollars, which facilitates the comparison of one selection model with another.

E. THEORY OF UTILITY ANALYSIS

In the context of utility analysis, there are four outcomes of interest associated with selection decisions. These outcomes are:

- Valid Positives (VP), which refers to the number of applicants that are hired and who turn out to be successful on the job.
- False Positives (FP), which refers to the number of applicants that are hired and who turn out to be unsuccessful on the job.
- False Negatives (FN) are the people who were not hired, but who would have been successful if they had been hired.
- Valid Negatives (VN) are the people that were not hired, and who would have been unsuccessful if they had been hired.

It is obvious from the terminology and the explanations that VP and VN constitute correct selection decisions, and FP and FN represent selection error.

These outcomes are perhaps easier to understand with the aid of a diagram. Figure 5.1 shows the relationship between hypothetical predicted (from a model) and actual scores on a job performance criterion for a large number of job applicants.

The ellipse contains the data on predicted and actual criterion scores. In this diagrammatic example, the correlation between the predicted and actual scores (the model's validity) is apparent--higher predicted scores are associated with higher actual scores and vice versa. The point y on the vertical axis is the dividing line between what is considered to be successful performance (say completion of 48 months of service for first term enlistees), and unsuccessful performance (less than 48 months service before

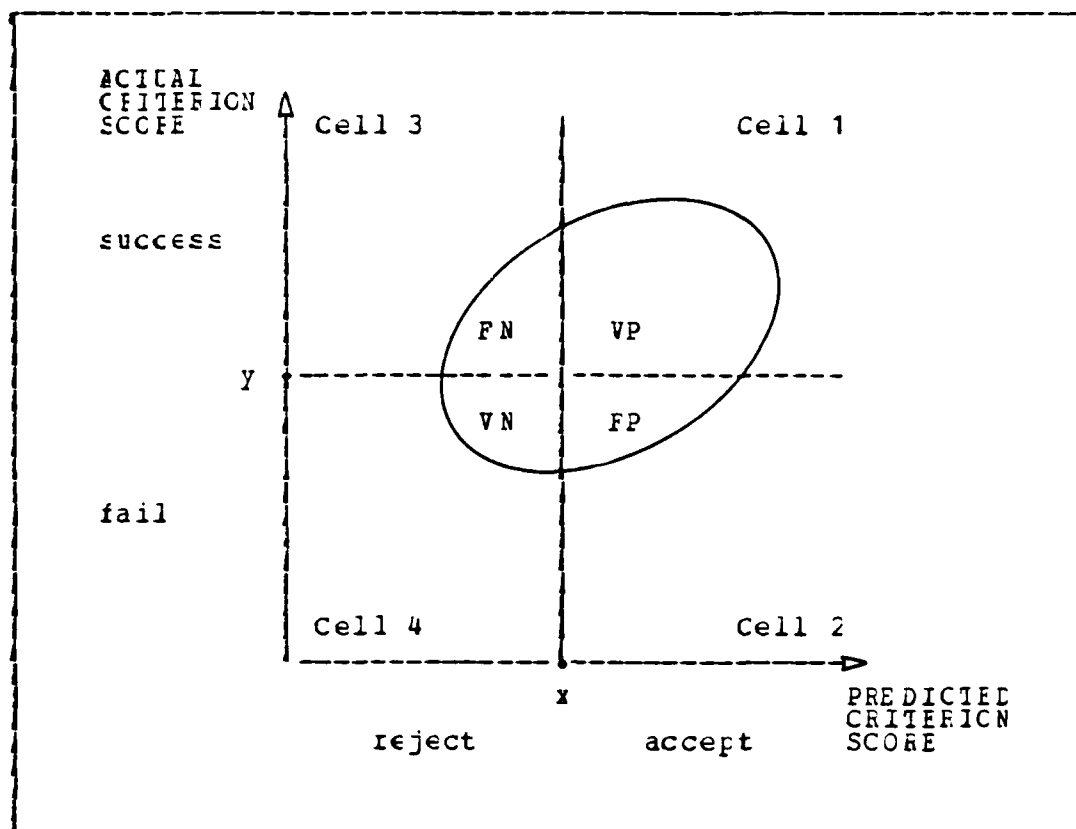


Figure 5.1 Hypothetical Predicted and Actual Scores

discharge). In utility analysis the term base rate is defined as that proportion of current employees who are considered to be successful. If seven out of every ten employees are successful, then the base rate is .70. The point x on the horizontal axis is referred to as the cut score. If an applicant's predicted score (from the model) is greater than x, then that person will be accepted (hired), and if their predicted score is less than x, then they will be rejected (not hired). The location of x on the horizontal axis will often depend on the selection ratio, which is the proportion of applicants that need to be accepted in order to fill a certain number of jobs. If,

over the course of one year, 80 job vacancies are expected to occur and if 100 applicants over the year are expected to apply for those jobs, then the selection ratio needs to be .80 if all vacancies are to be filled. In the happy circumstance (from the recruiter's point of view) where there are far more applicants than jobs, then the cut score x will be chosen so as to maximize the utility of the selection procedure. Utility is defined here to mean the expected gain in dollars that results from a particular selection strategy.

The lines generated from the base rate and the cut score divide the sample into four cells as shown. Each cell contains the people who are classified into each of the four outcomes of interest. In cells 1 and 2 are people whose predicted score is higher than the cut score. Therefore these people would be classified as accept. These accepted people (the positives) are further divided into those who would be successful (valid positives) and those who would be unsuccessful (false positives). Cells 3 and 4 contain the people who scored lower than the cut score on the predictor, and these would be classified as reject. Again, some of these rejected cases would have been successful (false negatives), and some would have failed (valid negatives). In utility analysis it is convenient to convert the cell counts (represented by VP, FP, FN and VN) to proportions of the overall sample, so each count is divided by the number of people in the sample and the cell probabilities (PVP, FFP, FFN and PVN respectively) result.

One further result of interest is the success rate. The success rate is defined as the proportion of hired applicants who are, or will be, successful. It is simply found by dividing FVP by the sum of FVP and FFP.

Given the concepts and terminology outlined above, it is now possible to discuss in general terms the factors that will affect the cell probabilities which, in turn, affect the expected utility.

1. Model Validity

The model's validity, as measured by the correlation between predicted and actual scores, is one factor that determines the degree of selection error resulting from the selection strategy. If the validity is high, then the proportions of correctly classified people (PVP and PVN) will be higher, and the selection error (FFP and PFN) will be lower. Vineberg and Joyner in their review of almost 150 military studies related to job performance prediction, found that validities range from .15 to .40, from a total of 350 validity coefficients [Ref. 15]. Generally, validities within this range would be considered as low or medium.

2. Base Rate

If the existing base rate is high (say .70 or greater), then it means that whatever selection strategy is currently in use has a high rate of success in identifying potentially successful applicants. Under these circumstances, it is unlikely that using a new model in the staffing process would yield much of an improvement in correctly selecting applicants. A high base rate means that the cell probabilities for FVP and PFN are going to be higher than for FFP and PVN.

3. Selection Ratio

Assuming the model is valid, the lower the selection ratio, the more useful the model will be in identifying successful applicants. Decreasing selection ratios mean that the organization can be increasingly selective in whom it hires. Naturally, it will tend to accept only those who score highest on the predictor, those who are also predicted most likely to be successful. A low selection ratio (high cut score) will mean that FVP and FFP will be small. It

also follows that a low selection ratio will yield a higher success rate--although few people will be hired, most of them will represent correct selection decisions (FVP).

C. ESTIMATING THE UTILITY OF A MODEL

The expected utility (EU) of a model is found by summing the products of each cell probability and its associated cell utility (U1, U2, U3 and U4), and subtracting the cost of giving the test to an applicant (UT).

$$EU = U1(FVP) + U2(FFP) + U3(FFN) + U4(PVN) - UT \quad (5.1)$$

Appendix E contains detailed descriptions on how cell probabilities and cell utilities are determined. For a discriminant model the cell probabilities may be readily derived from the output of the SAS Discriminant procedure, because the model classifies cases into predicted successes and predicted failures. In the regression model the cut score is not known in advance, so cell probabilities that result from a number of possible cut scores are calculated, and a cut score is eventually chosen based on which set of cell probabilities maximizes the utility of the model.

The formula for calculating the expected utility of a model requires that a utility be assigned to each selection outcome. These cell utilities are designated U1 through U4 and are associated with the outcomes VP, FP, FN and VN respectively. The Billet Cost Model provides an estimate of the cost to the Navy of staffing a billet. In this study it is assumed that this cost is equal to the marginal product of a successful sailor, and so the utility of a valid positive (U1) is assigned a value of \$24,163 [Ref. 16]. No proven technique exists for estimating the cell utilities for the three other selection outcomes. Individual

circumstances and prevailing market conditions make it difficult to estimate these outcomes with real confidence, so these cell utilities were estimated relative to U1, and a minor form of sensitivity analysis was conducted. The cell utility of a false positive (U2) was assigned values of -.5, -1 and -2. Valid negatives (U4) were assigned an equal and opposite utility to U2, and false negatives (U3) were assigned values of 0, -.25 and -.5. Table VIII shows seven different sets of cell utilities that were considered.

TABLE VIII
Relative Cell Utilities

U1	U2	U3	U4
1.0	-0.5	-0.25	0.5
1.0	-0.5	0.0	0.5
1.0	-1.0	0.0	1.0
1.0	-2.0	0.0	2.0
1.0	-0.5	-0.5	0.5
1.0	-1.0	-0.5	1.0
1.0	-2.0	-0.5	2.0

The cost of administering a test (UT) is of significance if the costs of testing are different for competing selection strategies. The models developed in this study use data gathered from the existing tests, and therefore the costs of testing will remain much the same. Thus in this context, UT may be ignored since it applies equally to the old and new tests.

E. RESULTS OF UTILITY ANALYSIS

Tables IX and X contain the results of the utility analysis of the regressor and discriminant models respectively. The "Percent Change in EU" column is the result of the comparison of the model's utility with the utility of the

Navy's original selection strategy (base line utility). A positive percentage change in EU indicates that the maximum utility obtainable from the model is higher than the utility of the original selection strategy. An increase in utility of say \$50 means that the Navy saves \$50 for each selection decision (correct or incorrect) that is made by using the model rather than the original strategy. For the models with the SUCCTAF or TAFMS1 criterion, the base rate is .861, i.e., 86.1 percent of the people selected by the Navy were successful. These people can be thought of as the valid positives of the original strategy and the remaining 13.9 percent are false positives. (For the SUCCESS2 criterion these figures are 76.8 percent and 23.3 percent.) Unfortunately it is not possible to calculate the values of false and valid negatives so these are considered to be zero. For the TAFMS1 or SUCCTAF criterion then, the cell probabilities for the original selection strategy are FVF = .861, PFF = .139, PFN = 0 and FVN = 0. The base line utility for each of the three different combinations of U1 and U2 can then be calculated. The model utilities are then compared to these base line utilities and the differences, expressed as a percentage of the base line utilities, are reported. Similarly the base success rate of the original strategy is also .861 (for the TAFMS1 or SUCCTAF criterion). The column "Change in Succrate" reports the actual difference between the models' success rates and the base success rates. The column "SRATIO" shows the selection ratio that results when the cut score is chosen so as to maximize the utility, for each set of cell utilities.

1. Regression Models

For most sets of cell utilities, the regression models developed show little improvement over the original selection strategy. In most cases the selection ratio is

very close to 1 and the percentage increase in expected utility is very small. This is not a surprising result because the model validities are relatively low (around .20) and, more significantly, the base rates are very high at .861 and .768. It is interesting to note, however, that when the costs of a false positive and the benefits of a valid negative are high, then the selection ratio is driven down, and the utility and success rate go up.

2. Discriminant Models

In general the discriminant models did not perform as well as the regression models or the Navy's original selection strategy. For some models the percent change in EU was a significant positive number, but these were usually associated with extreme assumptions of cell utilities. In addition to the factors mentioned in the previous subsection, this poor performance is because the discriminant models lack the flexibility to vary the cell probabilities depending on the values of the cell utilities. There is no option to vary predictions depending on the consequences of correct and incorrect selection decisions, and thus only one set of cell probabilities is available for each discriminant model.

TABLE IX
Utility Results - Regression Models

MODEL	U1	U2	U3	U4	Δ% EU	Δ SUCRATE	SFATIC
A	1.0	-0.5	-0.25	0.5	0.12	0.001	C.9993
	1.0	-0.5	0	0.5	0.14	0.001	C.9993
	1.0	-1.0	0	1.0	0.34	0.001	C.9993
	1.0	-2.0	0	2.0	5.85	0.022	C.9993
	1.0	-0.5	-0.5	0.5	0.11	0.001	C.9993
	1.0	-1.0	-0.5	1.0	0.32	0.001	C.9993
E	1.0	-2.0	-0.5	2.0	1.25	0.003	C.9993
	1.0	-0.5	-0.25	0.5	0.0	0.0	1.0
	1.0	-0.5	0	0.5	0.0	0.0	1.0
	1.0	-1.0	0	1.0	0.0	0.0	1.0
	1.0	-2.0	0	2.0	6.28	0.023	C.810
	1.0	-0.5	-0.5	0.5	0.0	0.0	1.0
C	1.0	-1.0	-0.5	1.0	0.0	0.0	1.0
	1.0	-2.0	-0.5	2.0	0.28	0.002	C.985
	1.0	-0.5	-0.25	0.5	0.0	0.0	1.0
	1.0	-0.5	0	0.5	0.0	0.0	1.0
	1.0	-1.0	0	1.0	0.05	0.001	C.998
	1.0	-2.0	0	2.0	5.79	0.016	C.971
D	1.0	-0.5	-0.5	0.5	0.0	0.0	1.0
	1.0	-1.0	-0.5	1.0	0.0	0.0	1.0
	1.0	-2.0	-0.5	2.0	0.4	0.004	C.972
	1.0	-0.5	-0.25	0.5	0.15	0.002	C.995
	1.0	-0.5	0	0.5	0.22	0.002	C.995
	1.0	-1.0	0	1.0	5.1	0.027	C.861
F	1.0	-2.0	0	2.0	72.98	0.074	C.283
	1.0	-0.5	-0.5	0.5	0.03	0.002	C.995
	1.0	-1.0	-0.5	1.0	0.76	0.002	C.995
	1.0	-2.0	-0.5	2.0	35.44	0.014	C.806
	1.0	-0.5	-0.25	0.5	0.0	0.0	1.0
	1.0	-0.5	0	0.5	0.0	0.0	1.0
E	1.0	-1.0	0	1.0	3.51	0.033	C.799
	1.0	-2.0	0	2.0	61.76	0.124	C.050
	1.0	-0.5	-0.5	0.5	0.0	0.0	1.0
	1.0	-1.0	-0.5	1.0	0.0	0.0	1.0
	1.0	-2.0	-0.5	2.0	33.51	0.033	C.799
	1.0	-0.5	-0.25	0.5	0.14	0.001	C.997
F	1.0	-0.5	0	0.5	0.16	0.001	C.997
	1.0	-1.0	0	1.0	4.77	0.013	C.816
	1.0	-2.0	0	2.0	79.18	0.063	C.511
	1.0	-0.5	-0.5	0.5	0.11	0.001	C.997
	1.0	-1.0	-0.5	1.0	0.46	0.001	C.997
	1.0	-2.0	-0.5	2.0	36.61	0.014	C.807

Note: The base utilities for Models A, E and C are - \$19112 (when U2 is -0.5), \$17428 (when U2 is -1.0) and \$14061 (when U2 is -2), and the base success rate is 0.861.

The base utilities for Models D, E and F are - \$15744 (when U2 is -0.5), \$12938 (when U2 is -1.0) and \$7326 (when U2 is -2.0), and the base success rate is 0.768.

TABLE X
Utility Results - Discriminant Models

MODEL	U1	U2	U3	U4	Δ% EU	Δ SUCCESS RATE	STATIC
A	1.0	-0.5	-0.25	0.5	-13.0	0.016	0.856
	1.0	-0.5	0	0.5	-9.5		
	1.0	-1.0	0	1.0	-5.8		
	1.0	-2.0	0	2.0	-4.6		
	1.0	-0.5	-0.5	0.5	-16.5		
	1.0	-1.0	-0.5	1.0	-13.3		
	1.0	-2.0	-0.5	2.0	-4.8		
B	1.0	-0.5	-0.25	0.5	-0.1	0.0	0.999
	1.0	-0.5	0	0.5	-0.1		
	1.0	-1.0	0	1.0	-0.1		
	1.0	-2.0	0	2.0	-0.1		
	1.0	-0.5	-0.5	0.5	-0.2		
	1.0	-1.0	-0.5	1.0	-0.1		
	1.0	-2.0	-0.5	2.0	0.0		
C	1.0	-0.5	-0.25	0.5	-5.0	0.074	0.954
	1.0	-0.5	0	0.5	-3.8		
	1.0	-1.0	0	1.0	-3.1		
	1.0	-2.0	0	2.0	-1.0		
	1.0	-0.5	-0.5	0.5	-6.2		
	1.0	-1.0	-0.5	1.0	-5.7		
	1.0	-2.0	-0.5	2.0	-4.3		
D	1.0	-0.5	-0.25	0.5	-86.8	0.081	0.178
	1.0	-0.5	0	0.5	-63.1		
	1.0	-1.0	0	1.0	-38.5		
	1.0	-2.0	0	2.0	-67.5		
	1.0	-0.5	-0.5	0.5	-110.0		
	1.0	-1.0	-0.5	1.0	-96.8		
	1.0	-2.0	-0.5	2.0	-25.3		
E	1.0	-0.5	-0.25	0.5	-13.3	0.033	0.799
	1.0	-0.5	0	0.5	-8.4		
	1.0	-1.0	0	1.0	-3.5		
	1.0	-2.0	0	2.0	-4.6		
	1.0	-0.5	-0.5	0.5	-18.2		
	1.0	-1.0	-0.5	1.0	-8.4		
	1.0	-2.0	-0.5	2.0	-33.5		
F	1.0	-0.5	-0.25	0.5	-53.1	0.069	0.432
	1.0	-0.5	0	0.5	-37.5		
	1.0	-1.0	0	1.0	-15.4		
	1.0	-2.0	0	2.0	-79.4		
	1.0	-0.5	-0.5	0.5	-68.7		
	1.0	-1.0	-0.5	1.0	-53.3		
	1.0	-2.0	-0.5	2.0	-12.5		

Note: The base utilities for Models A, B and C are - \$19112 (when U2 is -0.5), \$17428 (when U2 is -1.0) and \$14061 (when U2 is -2.0), and the base success rate is 0.861.

The base utilities for Models D, E and F are - \$15744 (when U2 is -0.5), \$12938 (when U2 is -1.0) and \$7326 (when U2 is -2.0), and the base success rate is 0.768.

VI. CONCLUSIONS AND RECOMMENDATIONS

This study set out to provide a method for developing enlistment standards models which improves upon similar processes presently in use. Toward that end, significant advances have been made, particularly when compared to prior studies conducted at the Naval Postgraduate School. The techniques used provide a much more comprehensive approach to model development. They employ regression analysis to fully develop the stepwise regression results. In addition, stepwise discriminant procedures were used to find an optimal model prior to full discriminant analysis. Alternative criteria for measuring successful operational performance, including a continuous length of service criterion, were incorporated in the models. Finally, each model was analyzed using both regression and discriminant analysis techniques.

Perhaps most significant is the presentation of a means by which the benefits from such efforts may be gauged. The development of innovative utility analysis programs affords future researchers an excellent opportunity to measure in monetary terms the benefits to be derived from implementing a new selection strategy. It is important to reiterate that the statistical and utility analysis techniques presented in this study may be easily applied or modified to accommodate selection standards model development for any of the more than 90 Navy ratings contained in the master data base.

A secondary purpose of this study was to discover whether the models developed improve upon existing selection and assignment strategy for the AD rating. By and large, the models presented do not appreciably enhance the processes used since 1976. The models do, however, allow

one to focus on some specific considerations in the current screening processes. For example, Models A, B, and C allow policy makers to consider length of service in months, and to vary the criterion for measuring success. This capability is particularly appropriate for use in a dynamic recruiting market.

A. RESULTS

This study yielded several other results worth noting. The term of enlistment variable may be used to predict success now that it has been corrected to reflect active duty obligation. This is particularly important when assessing Naval Reservists, whose six year contract generally requires only three years of active service. The change from a negative to a significantly positive correlation of TERMENLT on the criteria is one of the more important discoveries of this research effort.

This study also determined that the usefulness of the SCREEN composite score in predicting job performance measures was virtually nonexistent. It appears to be more appropriate to use the SCREEN score components in the models, at least when attempting to predict operational job performance. Nontraditional ASVAB subtest scores, such as Auto Information, may also be appropriate for use in the screening process. Another significant finding of this study is the definite presence of interaction effects. Considering personal measures on an individual in conjunction with other measures represents a marked change in current selection practices.

To summarize the results of the statistical analysis, the variables measuring term of enlistment, education, dependents status, sex and race emerged as repeatedly significant predictors of successful operational

performance. The composite measure of eligibility for the AD rating, and the ASVAB Auto Information subtest score, were also significant predictor variables. In addition, Model F was shown to be the best regression and discriminant model.

The results of the application of utility analysis show that the regression models developed in this study perform as well as or better than the original Navy strategy which was used as the comparison (base line utility). It is important to note however, that the methodology used in this part of the study ensured that regression models will provide a maximum utility at least equal to the base line utility. This is because the technique allows the cut score to be set so low that all cases are accepted. Models A and F are considered to be the best of the models because they provide for significant increases in utility without having to resort to impractically low selection ratios. The discriminant Models A and E are better than the others because improvement over base utility is possible, depending on the cell utilities.

As was mentioned in Chapter V, the high existing base rates are an indication that newly developed models are unlikely to produce superior results. Utility analysis is hindered by the difficulty of confidently estimating the individual cell utilities, and this is an area that is in need of further research. It is also difficult to compare new selection strategies to existing ones because it is impossible to classify the cases rejected by the existing strategy as valid or false negatives. Data of this sort can only be obtained by testing all applicants and then accepting all of them, regardless of their relationship to the cut score, or to the desired selection ratio.

E. RECOMMENDATIONS

Despite the advances made by this study, there remains many opportunities to refine the models presented for the AD rating, and to develop models for other Navy ratings. Procedurally, these opportunities include testing for curvilinearity of the models, expanding the interaction terms to three or more levels, and seeking different combinations of ASVAB subtest scores as potential predictors. There may also be other measures not evaluated by this study that are significant operational performance predictors, such as enlistment waivers, IEP status, or involvement with civil authorities.

Consideration should also be given to altering the criterion variables. One particularly promising adjustment may be to change the criterion to reflect achieving E-5. This may be appropriate since the models developed appear to do a better job of predicting longer LOS, as indicated by preliminary residual analysis. Developing separate models that yield predictions of shorter LOS may also be in order.

The multiple-stage analytic approach referred to in Chapter IV also appears to be a promising technique. Such analysis might consider change in dependent status, performance evaluations, or advancement exam results as variables in a model.

To improve the usefulness of utility analysis it is important that a technique be developed to estimate cell utilities with reasonable accuracy. Such a technique needs to be able to control for changes in the recruiting market, and be sensitive to the changing Navy requirements for recruits. It is also important that data be gathered on applicants who are not accepted into a particular rating, to allow researchers to determine if they were reclassified to another rating, or rejected entirely.

In conclusion, it is clear that continued efforts to develop selection standards models for all ratings are essential. For it is through these efforts that the cost of training and maintaining Navy personnel will be reduced. The resultant experienced career force will ensure the Navy is ready to meet any global commitment.

APPENDIX A

DATA BASE DEVELOPMENT PROGRAMS

This appendix provides the SAS programs used in this study to access the master data base, develop the AD data set, and create new predictor and criterion variables, as discussed in Chapter III. Each program contains the job control language information appropriate to the Naval Postgraduate School's IBM 3033 computer system. Statistical Analysis System (SAS) statements are employed in the programs to accomplish the desired functions. These SAS statements are normally preceded by comments to explain their purpose, the comments being identified by an asterisk.

Table XI contains the program called "ADSETUP". This program was used to access the master file and extract information on Aviation Machinist's Mates (AD). (The master file tape, originally called "ENLIST", has recently been revised and relabeled "NPS709".) The data file created by this program is called "ADDATA", and it contains the initial 243 variables from the master file. Also provided in the program are the variable names and labels. The program may be used to extract data from the master file for any of approximately 90 Navy ratings simply by entering the appropriate abbreviation and four digit code for the selected rating.

Table XII provides the program called "ADSCREEN" that was used to screen the data extracted from the master file. These screens were performed on observations in the "ADDATA" file, and the results were placed in a file called "ADSCREEN1". Because of the large number of cases and variables in the data, sufficient computing work space was not available. Therefore, the SAS KEEP statement was used to

retain 116 of the initial variables for analysis. It was felt these 116 variables captured all the desired measures on the observations that would be required for analysis. The last screen was incorporated following frequency distribution analysis to remove cases that had aberrant or impossible data associated with them.

Table XIII contains the program called "ADNEWVAR". This program was employed to create new predictor and criterion variables, as discussed in Chapter III. The program used information on observations in the "ADSUBSET" data file to create the new variables, and placed the results of these operations in a file called "ADALL4". This file thus constitutes the AD data set referred to throughout this study. It contains all of the selected and created variables that provide information on the 2820 ADs who remained in the data set after all screens were accomplished. It is this file that was used to conduct the statistical analysis for this study.

The "ADNEWVAR" program lists all created variable names and labels. It also contains the SAS statements that converted several qualitative variables to numeric variables or dichotomous (0,1) variables. Finally, the program shows the SAS statement used to split the AD data set into the two uniformly distributed random samples (RANDALL1). These derivation and validation samples were used during regression and discriminant model development described in Chapter IV.

TABLE XI

Program to Extract Data from the Master File

```

//ADDATA JCE (2807,0110), 'D OSIUND, SMC 1763', CLASS=K
//*MAIN CRG=NP GVM1.28C7P
//EXEC SAS
//SAS.WCFK DD SPACE=(CYL,(12,4))
//FILEIN DD UNIT=34CC-5,VOI=SER=ENLIST,
//DISF=CIL,DSN=ENLIST. ALL. A7678
//FILEOUT DD UNIT=333OV,MSVGP=FUB4B,DISP=(NEW,CATLG,DELETE),
//DSN=MSS.S28(7.ADDATA,
//LCB=(BLKSIZE=6400)
//SYSIN DD *
CPTICNS IS=80 NOCENTIF;

```

DATA FILEOUT.ADDATA;

* THIS PROGRAM EXTRACTS NEARLY ALL THE VARIABLES FROM THE MASTER FILE, AND WRITES OUT A FILE TO MASS STORAGE WHICH CONTAINS ALL THESE VARIABLES FOR ALL CASES WHICH HAD ANYTHING TO DO WITH THE 'AD' RATING.;

INFILE FILEIN;
INPUT

@ 5	CENSUSRG	PIE1.	@ 6	CENSUSDS	PIB1.	@ 7	HOMEZIE	FIB3.
@ 10	EMESTATE	PIE1.	@ 11	DATEDETY	PIB1.	@ 12	LATEDETM	FIB1.
@ 13	EFTHYR	PIE1.	@ 14	BIRTHMTH	PIB1.	@ 15	BIRTHDAY	PIB1.
@ 16	INTFYAGE	PIE1.	@ 17	RECCORCID	PIB1.	@ 18	HYEC	FIE1.
@ 19	SEX	PIE1.	@ 20	RACE	PIB1.	@ 21	ETHNIC	FIB1.
@ 22	BACFETEN	PIE1.	@ 23	MRTLDEND	PIB1.	@ 24	TESTFCRM	FIB1.
@ 25	AFCTFCNT	PIE1.	@ 26	AFCTGFPS	PIB1.	@ 27	ASVAEGI	PIB1.
@ 28	ASVAENC	PIE1.	@ 29	ASVAEAD	PIB1.	@ 30	ASVABWK	PIB1.
@ 31	ASVAEAR	PIE1.	@ 32	ASVABSP	PIB1.	@ 33	ASVAEMK	FIE1.
@ 34	ASVABEI	PIE1.	@ 35	ASVAEMC	PIB1.	@ 36	ASVABGS	PIE1.
@ 37	ASVAESI	PIE1.	@ 38	ASVABAI	PIB1.	@ 39	SERVACCS	FIB1.
@ 40	EBICESRV	PIE1.	@ 41	PUL	PIB1.	@ 42	HES	FIE1.
@ 43	ASVABCM	PIE1.	@ 44	ASVAECA	PIB1.	@ 45	ASVAECA	PIE1.
@ 46	ASVAECC	PIE1.	@ 47	ENTRYSTA	PIB1.	@ 48	HEIGHT	FIE1.
@ 49	WEIGHT	PIE1.	@ 50	SYSTICIBP	PIB1.	@ 51	DIATLBE	FIB1.
@ 52	MEDFAIL1	PIE1.	@ 53	MEDFAIL2	PIB1.	@ 54	MEDFAIL3	PIE1.
@ 55	WAIVER	PIE1.	@ 56	WAIVERAL	PIB1.	@ 57	EXAMSTAT	FIE1.
@ 58	INTFYHR	PIE1.	@ 61	TEEMENLT	PIB1.	@ 62	ENTRPAYG	PIB1.
@ 59	INTFYMTH	PIE1.	@ 60	ENTRYDAY	PIB1.			
@ 63	ECMECNTRY	PIE2.	@ 65	PROGENLT	PIB5.	@ 72	AFEESSTA	PIB1.
@ 73	ECNUSCFT	PIE1.	@ 74	ENISTCPT	PIB1.	@ 75	YOUTHFRG	FIB1.
@ 78	TAFILATE	PIE1.	@ 81	TRENIMOS	PIB5.	@ 86	TAFMS1	FIE2.
@ 88	DFCC1	PIE2.	@ 90	DDCC1	PIB2.	@ 92	HYEC1	PIB1.
@ 93	FAYGRDE1	PIE1.	@ 94	SERVICE1	PIB1.	@ 95	MRTSTAT1	FIE1.
@ 96	NDEFNNT1	PIE1.	@ 97	SPNSFL1	PIB3.	@ 100	ISC1	FIB1.
@ 101	SEFET1YR	PIE1.	@ 102	SEFET1MT	PIB1.	@ 103	SEPRT1DY	PIE1.
@ 104	EASCT1YR	PIE1.	@ 105	BASD1MTH	PIB1.	@ 106	BASD1DAY	PIB1.
@ 107	FJS1YR	PIE1.	@ 108	ETS1MNT	PIB1.			
@ 109	LCLE1YR	PIE1.	@ 110	DOLE1MTH	PIB1.			
@ 113	FEED1YR	PIE1.	@ 114	PEED1MTH	PIB1.	@ 115	FEED1DAY	FIE1.
@ 111	CHAFSEV1	PIE1.	@ 112	ELGAEUP1	PIB1.			
@ 116	FILEFLG1	PIE2.	@ 118	TAFMS2	PIB2.			
@ 120	DFCC2	PIE2.	@ 122	DDCC2	PIB2.	@ 124	HYEC2	PIB1.
@ 125	FAYGRDE2	PIE1.	@ 126	SERVICE2	PIB1.	@ 127	MRTSTAT2	PIE1.
@ 128	NDEFNNT2	PIE1.	@ 129	SPNSPD2	PIB3.	@ 132	ISC2	FIE1.
@ 133	SEFET2YR	PIE1.	@ 134	SEFET2MT	PIB1.	@ 135	SEPRT2DY	PIE1.
@ 136	EASCT2YR	PIE1.	@ 137	BASD2MTH	PIB1.	@ 138	BASD2DAY	PIE1.
@ 139	FTS2YR	PIE1.	@ 140	ETS2MNT	PIB1.			
@ 141	LCLE2YR	PIE1.	@ 142	DOLE2MTH	PIB1.			
@ 145	FEED2YR	PIE1.	@ 146	PEED2MTH	PIB1.	@ 147	FEED2DAY	FIE1.
@ 143	CHAFSEV2	PIE1.	@ 144	ELGAEUP2	PIB1.			

a148	FILEFIG2	PIE2.	a150	TA FMS3	PIB1.			
a151	TA FMS4	PIE1.	a152	DPOCC3	PIB2.	@154	DDOC 3	PIE2.
a156	EYEC3	PIE1.	a157	PAYGRLE3	PIB1.	@158	SERVICE3	PIB1.
a159	MTSTTA13	PIE1.	a160	NDENDNT3	PIB1.	@161	SPNSPC3	PIE3.
a165	SEFFET3YR	PIE1.	a166	SEPRT3MT	PIB1.	@167	SEPRT3DY	PIE1.
a168	EASL3YR	PIE1.	a169	BA SD3MTH	PIB1.	@170	EASD3DAY	PIE1.
a171	EIS3YEAR	PIE1.	a172	ET S3MTH	PIB1.			
a173	ECLE3YR	PIE1.	a174	DOLE3MTH	PIB1.			
a177	FEEL3YR	PIE1.	a178	PEED3MTH	PIB1.	@179	FEBD3DAY	PIB1.
a164	ISCS	PIE1.						
a175	CHAFSEV3	PIE1.	a180	FILEFIG3	PIB2.			
a176	ELGREUF3	PIB1.	a186	DOFYALEP	PIB1.	@187	DOEMTDEF	PIB1.
a182	FILEMTCH	PIE4.	a189	SPFLGML	PIB1.			
a188	MNTESDEP	PIE1.	a191	DCFGMNT	PIB1.			
a190	ICFGYR	PIE1.						
a212	GCT	2.	a214	ARI	2.	@216	MECH	2.
a218	CIEF	2.	a220	AFQTS	2.	@222	FNEC	\$4.
a227	CTZNSHIP	\$1.	a230	SECDEFND	\$1.	@231	ERCL	\$2.
a229	FRILEFND	\$1.	a234	AUTHRAIE	\$4.	@240	EDPGYR	\$4.
a233	GFUCFIND	\$1.	a245	SCHLWVR	\$1.	@246	ASTAR	\$1.
a244	SCHICCLE	\$1.	a250	PRESRATE	\$4.			
a247	TSSIND	\$1.	a255	PRETAERV	\$3.	@258	EXAMRATE	\$4.
a254	NUMEG1	\$1.	a263	EXRTAERV	\$3.	@266	ICTLRAW	\$3.
a262	NUMFG2	\$1.	a272	PRCODE	\$2.	@274	ALTPRCDE	\$2.
a269	SILNAVY	2.	a281	FNMTCUT	5.	@287	PRFFACTR	3.
a276	FINIMUIT	5.	a292	CHNGRATE	\$1.	@298	TYPENLST	\$2.
a290	AWIFACTR	2.	a297	SPERCIND	\$1.			
a296	EATEIND	\$1.	a302	NENLSMT	1.	@313	CAS	\$4.
a301	MCDEST	\$1.	a309	TAS	\$4.	@319	SIPG	\$4.
a303	FACS	YMMDD6.	a318	LOSWE	\$1.			
a317	ICSCCDE	\$1.	a324	TIR	\$4.	@349	DTIS	3.
a323	TIRWVR	\$1.	a343	EDFG	YMMDD6.	@384	AGE	2.
a336	ADEL	YMMDD6.	a356	NCHANGES	3.	@390	MENTLGRF	\$1.
a352	RECFORRES	1.	a388	NHRCAFCT	2.	@394	HYNDPND1	2.
a386	NHRCGCT	2.	a392	MOBLDSGN	\$1.	@399	REGRESRV	\$1.
a391	EDCERTIF	\$1.	a398	SSDUIY	\$1.	@402	SSNCENGE	\$1.
a396	GEF4PERCG	\$2.	a401	NOTRCMD	\$1.	@406	TOTLAWOL	1.
a400	EYFAYGFD	\$1.	a405	TOTLDEMO	1.	@409	TOTCVLCN	1.
a403	TCTEFCMO	2.	a408	TOTMLTCN	1.	@418	ATTRITCI	\$1.
a407	TCTLESRT	1.	a416	SCREEN	2.	@422	RECFRCGM	\$1.
a412	INGTHSFV	\$4.	a420	RECENIST	\$2.	@435	ELSTHIST	\$1.
a419	FECNTC	\$1.	a425	RCFGSCRT	\$4.	@444	NDAYSE4	4.
a423	FECFRGSC	\$2.	a440	NDAYSE3	4.	@456	DMDCUIC	\$6.
a436	NDAYSE2	4.	a452	DMDCNEC	\$4.	@468	GRAEDATE	YMMDD6.
a449	LMDCRATE	\$3.						
a462	CCNVLAIE	YMMDD6.						
a474	TRANLATE	YMMDD6.						
a480	FABNNEC	\$4.	a484	TRAININD	\$1.	@485	STACIION	\$1.;

IABEI
 CENSUSRG=CENSUS REGION (10 CODES)
 CENSUSIS=CENSUS DISTRICT (5 CODES)
 HCMEZIE=HOME OF RECCED ZIP CODE
 HMESTATE=HOME OF RECCFL--STATE
 LATELEFY=YEAR OF FINAL QUALIFYING DETERMINATION
 LATELEFE=MCNTH OF FINAL QUALIFYING DETERMINATION
 EIRTHYR=YEAR OF BIRTH
 EIRTHMTH=MCNTH OF BIRTH
 EIRTHDAY=DAY OF BIRTH
 ENTRYAGE=AGE OF INDIVIDUAL AT TIME OF ENTRY
 RECCFLID=RECORD ID--EXAM SCRE, DEP, ACTIVE DUTY
 EYEC=HIGHEST YEAR OF EDUCATION
 SEX=(1) MALE, (2) FEMALE
 RACE=(1) WHITE, (2) BLACK, (3) OTHER
 ETHNIC=INDIVIDUAL'S REPORTED ETHNIC STATUS
 RACEFTN=SIX RACE-ETHNIC COMBINATIONS
 MRTLDENC=MARITAL STATUS/DEPENDENTS
 TESTFCRM=TEST FORM/ECFA, ASVAB, AFWST, AFQT, OSP...
 AFQTECNT=AFQT PERCENTILE (OR EQUIVALENT)

AFOTGRES=AFCT GROUPS (5, 4C, 4B, 4A, 3B, 3A, 2, 1)
 ASVAEFGI=ASVAB APTITUDE AREA SCORE--SUBSCALE GI
 ASVAEENC=ASVAB APTITUDE AREA SCORE--SUBSCALE NO
 ASVAEAT=ASVAB APTITUDE AREA SCORE--SUBSCALE AD
 ASVAEWK=ASVAB APTITUDE AREA SCORE--SUBSCALE WK
 ASVAEAB=ASVAB APTITUDE AREA SCORE--SUBSCALE AR
 ASVAESP=ASVAB APTITUDE AREA SCORE--SUBSCALE SP
 ASVAEMK=ASVAB APTITUDE AREA SCORE--SUBSCALE MK
 ASVAEEI=ASVAB APTITUDE AREA SCORE--SUBSCALE EI
 ASVAEMC=ASVAB APTITUDE AREA SCORE--SUBSCALE MC
 ASVAEGS=ASVAB APTITUDE AREA SCORE--SUBSCALE GS
 ASVAESI=ASVAB APTITUDE AREA SCORE--SUBSCALE SI
 ASVAEAI=ASVAB APTITUDE AREA SCORE--SUBSCALE AI
 SERVACCS=SERVICE OF ACCESSION (NAVY, 2)
 ERICRSFV=PRIOR SERVICE (NON-PRIOR SERVICE, 1)
 FUL=GEN. HEALTH, UPPER & LOWER EXTREMITIES
 EES=HEARING, VISION, PSYCHIATRIC WELL-BEING
 ASVAECM=ASVAB APTITUDE AREA SCORE--SUBSCALE CM
 ASVAECA=ASVAB APTITUDE AREA SCORE--SUBSCALE CA
 ASVAECE=ASVAB APTITUDE AREA SCORE--SUBSCALE CE
 ASVAECC=ASVAB APTITUDE AREA SCORE--SUBSCALE CC
 ENTRYSTA=ENTRY STATUS (1, DIRECT TO ACTIVE DUTY)
 FEIGHT=HEIGHT IN INCHES (FRACTIONS DROPPED)
 WEIGHT=WEIGHT IN POUNDS (FRACTIONS ROUNDED)
 SYSTICIEP=ELCCD PRESSURE--SYSTOLIC
 DIASTIEP=ELCCD PRESSURE--DIASTOLIC
 MEDFAI1=PRIMARY MEDICALLY DISQUALIFYING DEFECT
 MEDFAI2=SECONDARY MEDICALLY DISQUALIFYING DEFECT
 MEDFAI3=TERTIARY MEDICALLY DISQUALIFYING DEFECT
 WAIVER=PERMIT CODE FOR AN OTHERWISE INELIGIBLE
 WAIVEEAL=WAIVER APPROVAL LEVEL AND EXPLANATION
 EXAMSTAT=EXAMINATION STATUS (1, FULLY QUALIFIED)
 TERMENIT=TERM OF ENLISTMENT (NO. OF YEARS)
 ENTREPAYG=ENTRY PAY GRADE (E00--O11)
 HOMEENTY=HOME OF RECORD COUNTY--FIPS
 EROGENIT=PROGRAM ENLISTED FOR--SERVICE UNIQUE
 AFESSTA=MILITARY ENTRANCE PROCESSING STATIONS
 EONUSCPT=ECONUS OPTION, COMBAT OR NON-COMBAT
 ENLSTCPT=ENLISTMENT OPTION
 YCUTHERG=YOUTH & RESERVE TRAINING PROGRAMS
 TAPELATE=MONTH OF FILE ON WHICH RECORD SUBMITTED
 TRENIACS=CCCUP, SPECIAL/RATING CHOICE UPON ENTRY
 TAFMS1=MONTHS OF TCIL, ACTIVE FED. MILIT. SERV.
 DPOC1=D.C.D. PRIMARY OCCUPATION CODE
 DLOC1=D.C.D. DUTY OCCUPATION CODE
 HVEC1=HIGHEST YEAR OF EDUCATION
 PAYGRDE1=PAY GRADE AS-OF-DATE-OF-FILE/SEPARATION
 SERVICE1=SERVICE CODE (2, NAVY)
 MRTSTAT1=MARITAL STATUS (1, OTHER, 2, MARRIED)
 NDFNENIT1=NUMBER OF DEPENDENTS (1, NONE)
 SPNSFD1=SEPARATION PROGRAM DESIGNATOR
 ISC1=INTER-SERVICE SEPARATION CODE
 SEPR1YR=YEAR OF SEPARATION (2ND DMDC SECTION)
 SEPR1MT=MONTH OF SEPARATION (2ND DMDC SECTION)
 SEPR1DY=DAY OF SEPARATION (2ND DMDC SECTION)
 EASD1YR=YEAR OF ACTIVE DUTY EASE DATE
 EASD1MT=MONTH OF ACTIVE DUTY EASE DATE
 EASD1DY=DAY OF ACTIVE DUTY EASE DATE
 ETS1YR=ESTIMATED YEAR OF FULFILLED ACTIVE DUTY
 ETS1MT=ESTIMATED MONTH OF FULFILLED ACTIVE DUTY
 CHARSEV1=CHARACTER OF SERVICE
 ELGREUP1=REENLISTMENT ELIGIBILITY
 FEED1YR=YEAR OF PAY ENTRY EASE DATE
 FEED1MT=MONTH OF PAY ENTRY EASE DATE
 FEED1DY=DAY OF PAY ENTRY EASE DATE
 ENTRYYR=YEAR OF ENTRY TO ACTIVE/D.E.P.
 ENTRYMT=MONTH OF ENTRY TO ACTIVE/D.E.P.
 ENTRYDY=DAY OF ENTRY TO ACTIVE/D.E.P.
 SEPR1YR=YEAR OF SEPARATION (2ND DMDC SECTION)

SEPR11M1=MONTH OF SEPARATION (2ND DMDC SECTION)
 SEPR11Y1=YEAR OF SEPARATION (2ND DMDC SECTION)
 EASD1Y1=YEAR OF ACTIVE DUTY EASE DATE
 EASD1M1=MONTH OF ACTIVE DUTY EASE DATE
 EASD1Y1=YEAR OF ACTIVE DUTY EASE DATE
 ETS1Y1=ESTIMATED YEAR OF FULFILLED ACTIVE DUTY
 ETS1M1=ESTIMATED MONTH OF FULFILLED ACTIVE DUTY
 FEBD1Y1=YEAR OF PAY ENTRY EASE DATE
 FEBD1M1=MONTH OF PAY ENTRY EASE DATE
 FEBD1Y1=YEAR OF PAY ENTRY EASE DATE
 FILEFIG1=FILE FLAG NC. 1
 FEBD2Y1=YEAR OF PAY ENTRY EASE DATE
 FEBD2M1=MONTH OF PAY ENTRY EASE DATE
 SEPR12Y1=YEAR OF SEPARATION (3RD DMDC SECTION)
 SEPR12M1=MONTH OF SEPARATION (3RD DMDC SECTION)
 SEPR12Y1=YEAR OF SEPARATION (3RD DMDC SECTION)
 EASD2Y1=YEAR OF ACTIVE DUTY EASE DATE
 EASD2M1=MONTH OF ACTIVE DUTY EASE DATE
 EASD2Y1=YEAR OF ACTIVE DUTY EASE DATE
 ETS2Y1=ESTIMATED YEAR OF FULFILLED ACTIVE DUTY
 ETS2M1=ESTIMATED MONTH OF FULFILLED ACTIVE DUTY
 FEBD2Y1=YEAR OF PAY ENTRY EASE DATE
 FEBD2M1=MONTH OF PAY ENTRY EASE DATE
 FEBD2Y1=YEAR OF PAY ENTRY EASE DATE
 TAFMS2=MONTHS OF TCIL. ACTIVE FED. MILIT. SERV.
 IPOC2=D.C.D. PRIMARY OCCUPATION CODE
 IDOC2=D.C.D. DUTY OCCUPATION CODE
 HVEC2=HIGHEST YEAR OF EDUCATION
 PAYGRADE2=PAY GRADE AS-CF-DATE-CF-FILE/SEPARATION
 SERVICE2=SERVICE CODE (2, NAVY)
 MRTSTA2=MARITAL STATUS (1, CEER, 2, MARRIED)
 NDPCNT2=NUMBER OF DEPENDENTS (1, NONE)
 SPNSFL2=SEPARATION PROGRAM DESIGNATOR
 ISC2=INTER-SERVICE SEPARATION CODE
 CHARSRV2=CHARACTER OF SERVICE
 ELGREUP2=REENLISTMENT ELIGIBILITY
 FILEFIG2=FILE FLAG NC. 2
 FEBD3Y1=YEAR OF PAY ENTRY EASE DATE
 FEBD3M1=MONTH OF PAY ENTRY EASE DATE
 FEBD3Y1=YEAR OF PAY ENTRY EASE DATE
 SEPR13Y1=YEAR OF SEPARATION (4TH DMDC SECTION)
 SEPR13M1=MONTH OF SEPARATION (4TH DMDC SECTION)
 SEPR13Y1=YEAR OF SEPARATION (4TH DMDC SECTION)
 EASD3Y1=YEAR OF ACTIVE DUTY EASE DATE
 EASD3M1=MONTH OF ACTIVE DUTY EASE DATE
 EASD3Y1=YEAR OF ACTIVE DUTY EASE DATE
 ETS3Y1=ESTIMATED YEAR OF FULFILLED ACTIVE DUTY
 ETS3M1=ESTIMATED MONTH OF FULFILLED ACTIVE DUTY
 FEBD3Y1=YEAR OF PAY ENTRY EASE DATE
 FEBD3M1=MONTH OF PAY ENTRY EASE DATE
 FEBD3Y1=YEAR OF PAY ENTRY EASE DATE
 TAFMS3=MONTHS OF TCIL. ACTIVE FED. MILIT. SERV.
 TAFMS4=MONTHS OF TCIL. ACTIVE FED. MILIT. SERV.
 IPOC3=D.C.D. PRIMARY OCCUPATION CODE
 IDOC3=D.C.D. DUTY OCCUPATION CODE
 HVEC3=HIGHEST YEAR OF EDUCATION
 PAYGRADE3=PAY GRADE AS-CF-DATE-CF-FILE/SEPARATION
 SERVICE3=SERVICE CODE (2, NAVY)
 MRTSTA3=MARITAL STATUS (1, CEER, 2, MARRIED)
 NDPCNT3=NUMBER OF DEPENDENTS (1, NONE)
 SPNSFL3=SEPARATION PROGRAM DESIGNATOR
 ISC3=INTER-SERVICE SEPARATION CODE
 CHARSRV3=CHARACTER OF SERVICE
 ELGREUP3=REENLISTMENT ELIGIBILITY
 FILEFIG3=FILE FLAG NC. 3
 FILEMTC3=4-BYTE PRIMARY FILE MATCH INDICATORS
 DCEYFL3=DCE YEAR INIC D.E.F.
 DCEMTFL3=DCE MONTH INIC D.E.F.

MNTHSDEF=MONTHS IN D.F.P.
 SPFLGMI=SPANISH FLAG MASTER/LCSS
 LCPGMNTH=MONTH OF DCPG
 LCPGYR=YEAR OF DCPG
 GCT=EASIC BATTERY GCT
 ARI=EASIC BATTERY ARI
 MECH=EASIC BATTERY MECH
 CLER=EASIC BATTERY CLER
 FNEC=NAVY ENLISTED JOB CODE
 CITZNSHIP=CITIZENSHIP CODE
 ERCL=BRANCH/CLASS
 GROUFIN=GROUP INDICATOR
 AUTHRATE=AUTHORIZED RATE (AEBR.)
 EDPGYR=EFFECTIVE DATE OF PAY GRADE
 SCHICCD= SCHOOL CODE
 SCHLWVR=SCHOOL WAIVER
 PRESRATE=PRESENT RATE CODE
 ERRTAEFFV=PRESENT RATE (AEBR.)
 EXAMRATE=EXAMINATION RATE CODE
 EXRTAEFFV=EXAMINATION RATE (AEBR.)
 TOTLFAW=TOTAL RAW SCORE
 STDNAVY=STANDARDIZED NAVY SCORE
 PRCOLE=PROCESS CODE
 ALTPRCDE=ALTERNATE PROCESS CODE
 FINLMULT=CANDIDATE'S FINAL MULTIPLE
 FNMLTCUT=FINAL MULTIPLE CUT
 PRFFACTR=PERFORMANCE FACTOR
 AWIFACTR=AWI FACTOR
 CHNGRATE=CHANGE OF RATE INDICATOR
 NENISMT=NUMBER OF ENLISTMENTS
 IAOS=EXPIRATION OF ACTIVE OBLIGATED SERVICE
 TAS=TOTAL ACTIVE SERVICE
 CAS=CUMULATIVE ACTIVE SERVICE
 SIPG=SERVICE IN PAY GRADE
 IOSCCD=LENGTH OF SERVICE
 LOSWVR=LENGTH OF SERVICE WAIVER
 TIRWVR=TIME IN RATE WAIVER
 TIR=TIME IN RATE
 ADBD=ACTIVE DUTY BASE DATE
 EDPG=EFFECTIVE DATE OF PAY GRADE
 LTIS=DRILL TIME IN SERVICE
 NCHANGES=NUMBER OF CHANGES/ENTRIES IN NHRC FILE
 AGE=CANDIDATE'S CURRENT AGE
 NHRCGCT=NHRC FILE'S GENRL. CLASSIFICATION TEST
 NHRCACFI=NHRC FILE'S ARMED FORCES QUALIFY. TEST
 MENTIGRP=MENTAL GROUP CODE
 EDCERTIF=EDUCATION CERTIFICATE
 MOBLDSCN=MILITARY OBLIGATION DESIGNATOR
 HYNDCNDT=HIGHEST NUMBER OF PRIMARY DEPENDENTS
 GRP4FFCG=GROUP IV (100K) PROGRAM CODE
 SSDUTY=SFA-SHORE DUTY INDICATOR
 REGRESRV=REGULAR RESERVE INDICATOR
 HYPAYGRD=HIGHEST PAY GRADE
 NOTRCMD=NOT RECOMMENDED FOR RE-ENLISTMENT
 SSNCNGE=SOCIAL SECURITY/NAME CHANGE
 TCTPRCMT=TOTAL PROMOTIONS
 TOTLDEMC=TOTAL DEMOTIONS
 TOTLAWCI=TOTAL UA/AWCI
 TCTDESRT=TOTAL DESERTIONS
 TOTMILCN=TOTAL MILITARY CONFINEMENTS
 TOTCVICN=TOTAL CIVILIAN CONFINEMENTS
 INGTESRV=LENGTH OF SERVICE
 SCREEN=SCREEN SCORE
 ATTRITCN=ATTRITION INDICATOR
 RECNTC=RECRUIT NAVAL TRAINING COMMAND
 RECENST=RECRUIT TYPE ENLISTMENT
 RECPECGM=RECRUIT PROGRAM AT ENLISTMENT
 RECPEGSC=RECRUIT PROGRAM/SCHCCI
 RC2GSCRT=RECRUIT PROGRAM/SCHCCI RATE

ELSTHIST1=ENLISTED HISTORY STATUS
 NDAYSSE2=COMPUTED NUMBER OF DAYS TO E-2 RATING
 NDAYSSE3=COMPUTED NUMBER OF DAYS TO E-3 RATING
 NDAYSSE4=COMPUTED NUMBER OF DAYS TO E-4 RATING
 DOLE1YI=YEAR OF LATEST RE-ENLISTMENT
 DOLE1MTH=MONTH OF LATEST RE-ENLISTMENT
 DOLE2YI=YEAR OF LATEST RE-ENLISTMENT
 DOLE2MTH=MONTH OF LATEST RE-ENLISTMENT
 DOLE3YI=YEAR OF LATEST RE-ENLISTMENT
 DOLE3MTH=MONTH OF LATEST RE-ENLISTMENT
 IMDCRATE=FINAL RATING AS LISTED BY D.M.D.C.
 IMDCNEC=FINAL N.E.C. AS LISTED BY D.M.D.C.
 IMDCUIC=FINAL U.I.C. AS LISTED BY D.M.D.C.
 CONVDATE=CONVENING DATE FOR NITRAS COURSE
 GRADDATE=GRADUATION DATE FOR NITRAS COURSE
 TRANSDATE=TRANSACTION DATE FOR NITRAS RECORD
 EARNNEC=DID CANDIDATE EARN AN NEC?
 TRAININD=TRAINING INDICATOR
 STACTICN=STUDENT ACTION CODES (PASS, P, ETC.);

* THIS SCREEN SELECTS ONLY THOSE CASES WHICH HAD ANY
 AFFILIATION WITH THE 'AD' RATING. THAT IS, THOSE CASES
 WHICH ARE LISTED IN THE DMDC FILE AS PRESENTLY AD'S
 (PRETAERV) OR AS FINALLY AD'S (IMDCRATE), OR AS SIGNING
 UP FOR AD'S (RCPGSCFT), OR AS HAVING TAKEN THE AD
 RATING EXAMINATION (EXAMRATE).;

IF IMDCRATE='AD' OR PRETAERV='AD' OR
 RCPGSCFT='6200' OR EXAMRATE='6200';

* THIS NEXT SECTION OUTPUTS BASIC FREQUENCIES TO CHECK
 THAT THE RATING SPECIFIC DATA HAS BEEN WRITTEN INTO
 THE FILE IN MASS STORAGE.;

PROC IFEC DATA=FILECIT1.ADDATA;
 TABLES IMDCRATE PRETAERV RCPGSCFT EXAMRATE;
 TITLE CHECKOUT FREQUENCIES FROM THE FILE ADDATA.;

/*
 //

TABLE XII
Program to Screen the AD Data

```
//ADSCREEN JOB (2807,C110), 'D CSLUND, SMC 1763',CLASS=E
//*MAIN CPG=NPGVM1.2&C7P
//SYFC SAS
//SAS.WCFC LD SPACE=(CYL,(12,4))
//FILEIN DD DISP=SHR,DSN=MSS.S2807.ADDATA
//FILECUT DD UNIT=33&CV,MSVGF=FUB4A,
//DISP=(NEW,CATLG,DELETE),DSN=MSS.S2807.ADSUBSET,
//DCE=(BIKSIZE=640C)
//SYSIN DD *
CPTICNS IS=80 NOCENTEF;
```

* THIS PROGRAM REDUCES THE NUMBER OF CASES IN THE DATA SET BY SCREENING ON CERTAIN VARIABLES. THE INTENT OF THE SCREEN IS SUMMARIZED ABOVE THE APPROPRIATE SAS STATEMENTS;

```
DATA FILECUT.ADSUBSET;
SET FILEIN.ADDATA;
```

* THE NUMBER OF VARIABLES IN THE DATA IS REDUCED TO REDUCE THE WORK SPACE REQUIREMENTS.;

```
KEEP
AFCTIGRES AFCTPCNT AGE ASVABAD ASVAEAI
ASVAEAB ASVAEBI ASVABGI ASVABGS ASVAEMC
ASVABMK ASVABNC ASVAESI ASVABSP ASVABWK
ATTRITCD AUTHRATE AWIFACTF BASD1DAY BASD1MTH
EAS11YE CHARSRV1 CHARSEV3 DDOC1 DDOC3
IMDCNEC DMECRATE DOLE1MTH DOLE1YR DFOC1
DECC3 EDCERTIF ELGREUP1 ELGREUP3 ENTRPAYG
ENTRYAGE ENTRYDAY ENTRYMTH ENTRYSTA ENTRYYR
ETHNIC ETS1MTH ETS1YEAR EXAMRATE EXRTABRV
FILEFLG1 FILEFLG3 FINLMULT FNMLTCUT HVEC
HYEC1 HYEC3 HYNDENDT HYPAYGRD ISC1
ISC3 LNGTHSEV MENTLGRP MOBLDSGN MRTLLCPND
MRTSTAT1 MRTSTAT3 NDAYSE2 NDAYSE3 NDAYSE4
NDENDNT1 NDENDNT3 NHRC AFCT NOTRCMD PAYGRDE1
PAYGRDE3 PEED1DAY PEBD1MTH PEBD1YR PRESRATE
ERFIACFR PRIORSEV PRRTAERV RACE RACEETHN
RCFGSCFT RECNLS1 RECOBID RECPRGSC REGRESRV
SCREEN SEPRT1LY SEPRT1MT SEPRT1YR SEPRT3DY
SEPRT3LY SEPRT3MT SEPRT3YR SERVACCS SERVICE1
SERVICE3 SEX SPNSFD1 SPNSPD3 SSNCHNGE
SIDNAVY TAFMS1 TAFMS3 TAFMS4 TERMENLT
TESTFORM TOTCVLCN TOTDESRT TOTLAWCL TCTLDEM
TCTIRAW TOTMLCN TCTPRCMC TRENLMOS WAIVER
WAIVERAL;
```

* THIS SCREEN SELECTS ONLY THOSE CASES WHOSE FINAL DMIC FATING IS AD.;

```
IF IMDCRATE EQ 'AD';
```

* THE FOLLOWING LINE SELECTS ONLY THOSE CASES WITH NO PRIOR SERVICE. TO FURTHER REMOVE POTENTIAL PRIOR SERVICE CASES THOSE WHO HAVE CHANGED THEIR SOCIAL SECURITY NUMBER ARE ALSO REMOVED FROM THE SAMPLE.;

```
IF ERICSEV=1; IF SSNCHNGE EQ 0;
```

* THE FOLLOWING STATEMENTS SELECT ONLY THOSE CASES WEC WERE TESTED ON ASVAB FORMS 5, 6 OR 7. ALSO THOSE CASES WITH PECULIARLY HIGH ASVAB SCORES ARE ELIMINATED FROM THE DATA SET.;

```
IF ((TESTFORM GE 35) AND (TESTFORM LE 37));
IF ASVABGI<=15; IF ASVABNC<=50; IF ASVABAD<=30;
IF ASVABAR<=20; IF ASVABSE<=20; IF ASVABMK<=20;
IF ASVABGS<=20; IF ASVABSI<=20; IF ASVABAI<=20;
IF ASVABWK<=30; IF ASVABEI<=30; IF ASVABMC<=20;
```

* THIS SCREEN ONLY KEEPS THOSE WHO SIGNED UP FOR NAVY OR NAVAL RESERVE.;

```
IF ((SERVACCS EQ 2) OR (SERVACCS EQ 8));
```

* ONLY THOSE CASES WEC WERE KNOWN TO HAVE SIGNED UP FOR AT LEAST FOUR YEARS ACTIVE DUTY ARE KEPT.;

```
IF REENLIST EQ '11';
```

* THE CASES ARE SCREENED TO INCLUDE ONLY THOSE WITH 'GOOD' OR 'BAD' INTERSERVICE SEPARATION CODES, 'GREY' CASES ARE ELIMINATED.;

```
IF ISC1=0 OR ISC1=1
  OR (ISC1 GE 60 AND ISC1 LT 90);
IF ISC3=0 OR ISC3=1
  OR (ISC3 GE 60 AND ISC3 LT 90);
```

* THIS NEXT SCREEN KEEPS THOSE CASES FOR WHICH CLEAR 'ELIGIBLE TO REENLIST' DATA IS AVAILABLE.;

```
IF ELGREUP1=0 OR ELGREUP1=1 OR ELGREUP1=4 OR
  (ELGREUP1=240 AND (ELGREUP3=0 OR ELGREUP3=1));
```

* THESE SCREENS ELIMINATE CASES WITH IMPOSSIBLE DATA.;

```
IF AFQTGRPS NE 0; IF ENTPAYG NE 0;
IF ICSEMONTHS LE 72; IF LENGTHSRV NE 0;
IF FACE NE 0; IF ETHNIC NE 0;
IF TAFMS1 LE 72; IF LENGTHSRV NE 0;
IF ENTRYAGE NE 17; IF AFQTPCNT NE 0;
IF FACEETHN NE 0;
```

/*
//

TABLE XIII

Program to Create New Variables

```
//ADNEWVAR JOB (2807,C110), 'D CSLUND, SMC 1763', CLASS=B
//*MAIN CRG=NPGVM1.28C7P
//EXEC SAS
//SAS.WORK LD SPACE=(CYL,(12,4))
//FILEIN DD DISP=SHR,DSN=MSS.S2807.ADSUBSET
//FILEOUT DD UNIT=333CV,MSVGP=FUE4A,
//DISP=(NEW,CATLG,DELETE),DSN=MSS.S2807.ADAIL4,
//LCE=(ELKSIZE=640C)
//SYSIN DD *
CPTICNS IS=80 NOCENTIF ;
```

* THE PURPOSE OF THIS PROGRAM IS TO GENERATE NEW VARIABLES FOR USE IN THE ANALYSIS, EITHER BY RECODING ORIGINAL VARIABLES, OR BY CREATING NEW VARIABLES;

```
DATA FILEOUT.ADAIL4;
SET FILEIN.ADSUBSET;
```

* THE FOLLOWING LINES CREATE DIFFERENT 'ENTRY GROUPS'.

	ENTRY	EXAM	DMDC
{1}	YES	YES	YES
{2}	YES	YES	NO
{3}	YES	NO	YES
{4}	YES	NO	NO
{5}	NO	YES	YES
{6}	NO	YES	NO
{7}	NO	NO	YES;

```
IF (RCPGSCRT='6200' AND EXAMRATE='6200' AND
DMDCRATE='AD') THEN ENTRYGRP=1;
IF (RCPGSCRT='6200' AND EXAMRATE='6200' AND
DMDCRATE NE 'AD') THEN ENTRYGRP=2;
IF (RCPGSCRT='6200' AND EXAMRATE NE '6200' AND
DMDCRATE='AD') THEN ENTRYGRP=3;
IF (RCPGSCRT='6200' AND EXAMRATE NE '6200' AND
DMDCRATE NE 'AD') THEN ENTRYGRP=4;
IF (RCPGSCRT NE '6200' AND EXAMRATE='6200' AND
DMDCRATE='AD') THEN ENTRYGRP=5;
IF (RCPGSCRT NE '6200' AND EXAMRATE='6200' AND
DMDCRATE NE 'AD') THEN ENTRYGRP=6;
IF (RCPGSCRT NE '6200' AND EXAMRATE NE '6200' AND
DMDCRATE='AD') THEN ENTRYGRP=7;
```

* IN THIS SECTION, THE DMDC VARIABLE 'HYEC' IS CONVERTED TO A CONTINUOUS VARIABLE REPRESENTING NUMBER OF YEARS OF EDUCATION;

```
IF HYEC=1 THEN CHYEC=3.5; IF HYEC=2 THEN CHYEC=8;
IF HYEC=3 THEN CHYEC=9; IF HYEC=4 THEN CHYEC=10;
IF HYEC=5 THEN CHYEC=11; IF HYEC=6 THEN CHYEC=12;
IF HYEC=7 THEN CHYEC=13; IF HYEC=8 THEN CHYEC=14;
IF HYEC=9 THEN CHYEC=15; IF HYEC=10 THEN CHYEC=16;
IF HYEC=11 THEN CHYEC=18; IF HYEC=12 THEN CHYEC=20;
IF HYEC=13 THEN CHYEC=11.5;
```

* A NEW CATEGORICAL VARIABLE 'HSDG' IS NOW CREATED. A HIGH SCHOOL GRADUATE IS CODED A '1' AND A NON HIGH SCHOOL GRADUATE OR A G.E.C. IS CODED '0'.

```
IF ((HYEC LE 5) OR (HYEC EQ 13)) THEN HSDG=0;
IF ((HYEC GE 6) AND (HYEC NE 13)) THEN HSDG=1;
```

[illegible]

IF	ASVABAI=5	THEN	SASVABAI=36	IF	ASVABAD=5	THEN	SASVABAD=26
IF	ASVABAI=6	THEN	SASVABAI=38	IF	ASVABAD=6	THEN	SASVABAD=29
IF	ASVABAI=7	THEN	SASVABAI=40	IF	ASVABAD=7	THEN	SASVABAD=31
IF	ASVABAI=8	THEN	SASVABAI=42	IF	ASVABAD=8	THEN	SASVABAD=34
IF	ASVABAI=9	THEN	SASVABAI=44	IF	ASVABAD=9	THEN	SASVABAD=36
IF	ASVABAI=10	THEN	SASVABAI=46	IF	ASVABAD=10	THEN	SASVABAD=39
IF	ASVABAI=11	THEN	SASVABAI=48	IF	ASVABAD=11	THEN	SASVABAD=41
IF	ASVABAI=12	THEN	SASVABAI=50	IF	ASVABAD=12	THEN	SASVABAD=44
IF	ASVABAI=13	THEN	SASVABAI=52	IF	ASVABAD=13	THEN	SASVABAD=46
IF	ASVABAI=14	THEN	SASVABAI=55	IF	ASVABAD=14	THEN	SASVABAD=49
IF	ASVABAI=15	THEN	SASVABAI=57	IF	ASVABAD=15	THEN	SASVABAD=51
IF	ASVABAI=16	THEN	SASVABAI=59	IF	ASVABAD=16	THEN	SASVABAD=54
IF	ASVABAI=17	THEN	SASVABAI=61	IF	ASVABAD=17	THEN	SASVABAD=57
IF	ASVABAI=18	THEN	SASVABAI=63	IF	ASVABAD=18	THEN	SASVABAD=59
IF	ASVABAI=19	THEN	SASVABAI=65	IF	ASVABAD=19	THEN	SASVABAD=62
IF	ASVABAI=20	THEN	SASVABAI=67	IF	ASVABAD=20	THEN	SASVABAD=64
IF	ASVABSI=0	THEN	SASVABSI=20	IF	ASVABAD=21	THEN	SASVABAD=67
IF	ASVABSI=1	THEN	SASVABSI=21	IF	ASVABAD=22	THEN	SASVABAD=69
IF	ASVABSI=2	THEN	SASVABSI=23	IF	ASVABAD=23	THEN	SASVABAD=72
IF	ASVABSI=3	THEN	SASVABSI=25	IF	ASVABAD=24	THEN	SASVABAD=74
IF	ASVABSI=4	THEN	SASVABSI=28	IF	ASVABAD=25	THEN	SASVABAD=77
IF	ASVABSI=5	THEN	SASVABSI=30	IF	ASVABAD=26	THEN	SASVABAD=79
IF	ASVABSI=6	THEN	SASVABSI=32	IF	ASVABAD=27	THEN	SASVABAD=80
IF	ASVABSI=7	THEN	SASVABSI=35	IF	ASVABAD=28	THEN	SASVABAD=80
IF	ASVABSI=8	THEN	SASVABSI=37	IF	ASVABAD=29	THEN	SASVABAD=80
IF	ASVABSI=9	THEN	SASVABSI=39	IF	ASVABAD=30	THEN	SASVABAD=80
IF	ASVABSI=10	THEN	SASVABSI=42	IF	ASVABEI=0	THEN	SASVABEI=20
IF	ASVABSI=11	THEN	SASVABSI=44	IF	ASVABEI=1	THEN	SASVABEI=20
IF	ASVABSI=12	THEN	SASVABSI=46	IF	ASVABEI=2	THEN	SASVABEI=21
IF	ASVABSI=13	THEN	SASVABSI=48	IF	ASVABEI=3	THEN	SASVABEI=22
IF	ASVABSI=14	THEN	SASVABSI=51	IF	ASVABEI=4	THEN	SASVABEI=24
IF	ASVABSI=15	THEN	SASVABSI=53	IF	ASVABEI=5	THEN	SASVABEI=26
IF	ASVABSI=16	THEN	SASVABSI=55	IF	ASVABEI=6	THEN	SASVABEI=27
IF	ASVABSI=17	THEN	SASVABSI=58	IF	ASVABEI=7	THEN	SASVABEI=29
IF	ASVABSI=18	THEN	SASVABSI=60	IF	ASVABEI=8	THEN	SASVABEI=31
IF	ASVABSI=19	THEN	SASVABSI=62	IF	ASVABEI=9	THEN	SASVABEI=32
IF	ASVABSI=20	THEN	SASVABSI=65	IF	ASVABEI=10	THEN	SASVABEI=34
IF	ASVABWK=0	THEN	SASVABWK=23	IF	ASVABEI=11	THEN	SASVABEI=36
IF	ASVABWK=1	THEN	SASVABWK=24	IF	ASVABEI=12	THEN	SASVABEI=37
IF	ASVABWK=2	THEN	SASVABWK=26	IF	ASVABEI=13	THEN	SASVABEI=39
IF	ASVABWK=3	THEN	SASVABWK=27	IF	ASVABEI=14	THEN	SASVABEI=41
IF	ASVABWK=4	THEN	SASVABWK=28	IF	ASVABEI=15	THEN	SASVABEI=42
IF	ASVABWK=5	THEN	SASVABWK=30	IF	ASVABEI=16	THEN	SASVABEI=44
IF	ASVABWK=6	THEN	SASVABWK=31	IF	ASVABEI=17	THEN	SASVABEI=46
IF	ASVABWK=7	THEN	SASVABWK=33	IF	ASVABEI=18	THEN	SASVABEI=48
IF	ASVABWK=8	THEN	SASVABWK=34	IF	ASVABEI=19	THEN	SASVABEI=49
IF	ASVABWK=9	THEN	SASVABWK=35	IF	ASVABEI=20	THEN	SASVABEI=51
IF	ASVABWK=10	THEN	SASVABWK=37	IF	ASVABEI=21	THEN	SASVABEI=53
IF	ASVABWK=11	THEN	SASVABWK=38	IF	ASVABEI=22	THEN	SASVABEI=54
IF	ASVABWK=12	THEN	SASVABWK=39	IF	ASVABEI=23	THEN	SASVABEI=56
IF	ASVABWK=13	THEN	SASVABWK=41	IF	ASVABEI=24	THEN	SASVABEI=58
IF	ASVABWK=14	THEN	SASVABWK=42	IF	ASVABEI=25	THEN	SASVABEI=59
IF	ASVABWK=15	THEN	SASVABWK=44	IF	ASVABEI=26	THEN	SASVABEI=61
IF	ASVABWK=16	THEN	SASVABWK=45	IF	ASVABEI=27	THEN	SASVABEI=62
IF	ASVABWK=17	THEN	SASVABWK=46	IF	ASVABEI=28	THEN	SASVABEI=64
IF	ASVABWK=18	THEN	SASVABWK=48	IF	ASVABEI=29	THEN	SASVABEI=66
IF	ASVABWK=19	THEN	SASVABWK=49	IF	ASVABEI=30	THEN	SASVABEI=68
IF	ASVABWK=20	THEN	SASVABWK=50	IF	ASVAENO=0	THEN	SASVAENO=20
IF	ASVABWK=21	THEN	SASVABWK=52	IF	ASVAENO=1	THEN	SASVAENO=20
IF	ASVABWK=22	THEN	SASVABWK=53	IF	ASVAENO=2	THEN	SASVAENO=21
IF	ASVABWK=23	THEN	SASVABWK=55	IF	ASVAENO=3	THEN	SASVAENO=22
IF	ASVABWK=24	THEN	SASVABWK=56	IF	ASVAENO=4	THEN	SASVAENO=23
IF	ASVABWK=25	THEN	SASVABWK=57	IF	ASVAENO=5	THEN	SASVAENO=24
IF	ASVABWK=26	THEN	SASVABWK=59	IF	ASVAENO=6	THEN	SASVAENO=25
IF	ASVABWK=27	THEN	SASVABWK=60	IF	ASVAENO=7	THEN	SASVAENO=26
IF	ASVABWK=28	THEN	SASVABWK=62	IF	ASVAENO=8	THEN	SASVAENO=27
IF	ASVABWK=29	THEN	SASVABWK=63	IF	ASVAENO=9	THEN	SASVAENO=28
IF	ASVABNCO=10	THEN	SASVABNCO=29	IF	ASVABNO=31	THEN	SASVABNO=50
IF	ASVABNCO=11	THEN	SASVABNCO=30	IF	ASVABNO=32	THEN	SASVABNO=51

```

IF ASVAENC=12 THEN SASVABNO=31; IF ASVABNO=33 THEN SASVAENC=52;
IF ASVAENC=13 THEN SASVABNO=32; IF ASVABNO=34 THEN SASVAENC=53;
IF ASVAENC=14 THEN SASVABNO=33; IF ASVABNO=35 THEN SASVAENC=54;
IF ASVAENC=15 THEN SASVABNO=34; IF ASVABNO=36 THEN SASVAENC=55;
IF ASVAENC=16 THEN SASVABNO=35; IF ASVABNO=37 THEN SASVAENC=56;
IF ASVAENC=17 THEN SASVABNO=36; IF ASVABNO=38 THEN SASVAENC=57;
IF ASVAENC=18 THEN SASVABNO=37; IF ASVABNO=39 THEN SASVAENC=58;
IF ASVAENC=19 THEN SASVABNO=38; IF ASVABNO=40 THEN SASVAENC=59;
IF ASVAENC=20 THEN SASVABNO=39; IF ASVABNO=41 THEN SASVAENC=60;
IF ASVAENC=21 THEN SASVABNO=40; IF ASVABNO=42 THEN SASVAENC=61;
IF ASVAENC=22 THEN SASVABNO=41; IF ASVABNO=43 THEN SASVAENC=62;
IF ASVAENC=23 THEN SASVABNO=42; IF ASVABNO=44 THEN SASVAENC=63;
IF ASVAENC=24 THEN SASVABNO=43; IF ASVABNO=45 THEN SASVAENC=64;
IF ASVAENC=25 THEN SASVABNO=44; IF ASVABNO=46 THEN SASVAENC=65;
IF ASVAENC=26 THEN SASVABNO=45; IF ASVABNO=47 THEN SASVAENC=66;
IF ASVAENC=27 THEN SASVABNO=46; IF ASVABNO=48 THEN SASVAENC=67;
IF ASVAENC=28 THEN SASVABNO=47; IF ASVABNO=49 THEN SASVAENC=68;
IF ASVAENC=29 THEN SASVABNO=48; IF ASVABNO=50 THEN SASVAENC=69;
IF ASVAENC=30 THEN SASVABNO=49;

```

* THE FOLLOWING STATEMENTS CREATE THE NUMERIC VARIABLE
'LCSMNTHS' FROM THE VARIABLE 'INTHSRV';

```

YEAR=SUESTE(LNGTHSRV,1,2);
MCNTH=SUESTE(LNGTHSRV,3,2);
YEARS=YEAR*0;
MCNTHS=MCNTH*0;
LCSMNTHS=YEARS*12+MCNTHS;

```

* RECCDING TO A CATEGORICAL VARIABLE.;

```

IF METLDEND=10 THEN DEPENDTS=0; ELSE DEPENDTS=1;

```

* CONVERTING CHARACTER VARIABLES TO NUMERIC.;

```

NUXEYFAY=EYPAYGRD*0; NUNCTRFC=NOTRCMD*0;

```

* TO DEFINE THE HIGHEST PAYGRADE ACHIEVED, ACCORDING
TO THE DMIC FILE.;

```

IF FILEFIG1=8209 THEN PAYGRADE=PAYGRDE1;
IF FILEFIG1=8209 THEN PAYGRADE=PAYGRDE3;
IF PAYGRADE=0 THEN PAYGRADE=PAYGRDE1;
IF PAYGRADE=0 THEN PAYGRADE='.';

```

* CREATING THE ASVAB COMPOSITE VARIABLE USED WHEN
CLASSIFYING AD'S, AND ASSIGNING A DUMMY VARIABLE
TO IDENTIFY THOSE WAC ACHIEVED THE MINIMUM SCORE.;

```

ADCCMFCS = SASVABAF+SASVABEI+SASVABGS+SASVABMK;
IF ADCCMFCS GE 190 THEN ADMINSCR=1;
ELSE ADMINSCR=0;

```

* SETTING UP DUMMY VARIABLES TO ALLOW ANALYSIS OF
RACE AND SEX EFFECTS.;

```

IF RACE=1 THEN WHITE=1; ELSE WHITE=0;
IF RACE=2 THEN BLACK=1; ELSE BLACK=0;
IF RACE=3 THEN OTHER=1; ELSE OTHER=0;
IF SEX=2 THEN NUSEX=0; ELSE NUSEX=1;

```

* CREATING A RANDOM VARIABLE TO ALLOW THE DATA TO
BE SPLIT RANDOMLY IN HALF;

```

IF RANDU(0) <= .5 THEN RANDALL1=1; ELSE RANDALL1=0;

```

* CREATING INTERACTION VARIABLES FOR USE IN THE MODEL DEVELOPMENT.;

```
INTERC1=DEPENDTS*HSDG;
INTERC2=DEPENDTS*ELACK;
INTERC3=DEPENDTS*NUSEX;
INTERC4=DEPENDTS*TERMENLT;
INTERC5=DEPENDTS*SASVABAI;
INTERC6=DEPENDTS*ADMINSCR;
INTERC7=HSDG*ELACK;
INTERC8=HSDG*NUSEX;
INTERC9=HSDG*TERMENLT;
INTERC10=HSDG*SASVABAI;
INTERC11=HSDG*ADMINSCR;
INTERC12=ELACK*NUSEX;
INTERC13=ELACK*TERMENLT;
INTERC14=ELACK*SASVABAI;
INTERC15=ELACK*ADMINSCR;
INTERC16=NUSEX*TERMENLT;
INTERC17=NUSEX*SASVABAI;
INTERC18=NUSEX*ADMINSCR;
INTERC19=TERMENLT*SASVABAI;
INTERC20=TERMENLT*ADMINSCR;
INTERC21=SASVABAI*ADMINSCR;
```

* THE FOLLOWING LINES CREATE DIFFERENT CRITERION VARIABLES.;

```
IF ((SERVICE1 EQ 2) AND ((PAYGRADE GE 4) AND
  (NUHPAY GE 4))) THEN SUCCPAYG=1;
  ELSE SUCCPAYG=0;
IF ENTRYYR=78 AND ENTRYMTH GE 10 THEN LATEENLT=1;
  ELSE LATEENLT=0;
IF TAFMS1 GE 48 OR (TAFMS1 GE 45 AND LATEENLT=1)
  THEN SUCC2AF=1; ELSE SUCC2AF=0;
IF ELGREUP1=4 THEN SUCCREUP=0; ELSE SUCCREUP=1;
IF SUCCREUP=1 AND SUCC2AF=1 AND SUCCPAYG=1
  THEN SUCCESS2=1; ELSE SUCCESS2=0;
```

```
IABEL
ESDG      =HIGH SCHOOL GRADUATE (1), OTHER (0)
DEPENDTS  =SINGLE, NO DEPENDENTS (0), OTHERWISE (1)
CHYEC     =CONVERTED NUMBER OF YEARS OF EDUCATION
NUHPAY    =NHFC FILE--HIGHEST PAYGRADE ATTAINED
NUNCTEC   =NHRC--NCT RECOMMENDED FOR RE-ENLISTMENT
PAYGRADE  =DMDC-BASED HIGHEST PAY-GRADE ATTAINED
SASVAEGI  =STANDARDIZED SCORE - GENERAL INFORMATION
SASVAENC  =STANDARDIZED SCORE - NUMERICAL OPERATIONS
SASVAEAL  =STANDARDIZED SCORE - ATTENTION TO DETAIL
SASVAEWK  =STANDARDIZED SCORE - WORD KNOWLEDGE
SASVAEAR  =STANDARDIZED SCORE - ARITHMETIC REASONING
SASVAESF  =STANDARDIZED SCORE - SPACE PERCEPTION
SASVAEMK  =STANDARDIZED SCORE - MATH KNOWLEDGE
SASVAEEI  =STANDARDIZED SCORE - ELECTRONIC INFO
SASVAEMC  =STANDARDIZED SCORE - MECH COMPREHENSION
SASVAEFS  =STANDARDIZED SCORE - GENERAL SCIENCE
SASVAESI  =STANDARDIZED SCORE - SHCP INFORMATION
SASVAEAI  =STANDARDIZED SCORE - AUTO INFORMATION
WHITE     = (1) WHITE, ELSE (0)
ELACK     = (1) BLACK, ELSE (0)
CTHER     = (1) NEITHER BLACK NOR WHITE, ELSE (0)
NUSEX     = (1) MALE, (0) FEMALE
ADCOMFCS  =AD ASVAE COMPOSITE
ADMINSCR  =AD ASVAE COMPOSITE SCREEN
RANLAIL1  =VAR. TO ALLOC A RANDOM 50-50 SPLIT
IOSMNTHS  =LENGTH OF SERVICE IN MONTHS
ENTRYGFE  =ENTRY GROUP CLASSIFICATIONS
LATEENLT  =ENTERED AFTER SEP 78 (1), OTHERWISE (0)
SUCC2AF   =SUCCESS ON ICS CRITERION (1)
```

SUCCFAYG= (1,0) SUCCESS ON PAYGRADE
 SUCCFUF= (1,0) ELIGIBLE TO REENLIST
 SUCCFSS2= SUCCESS ON COMPOSITE CRITERION (1)
 INTERC1= DEPENDTS*HSDG
 INTERC2= DEPENDTS*BLACK
 INTERC3= DEPENDTS*NUSEX
 INTERC4= DEPENDTS*TERMENLT
 INTERC5= DEPENDTS*SASVABAI
 INTERC6= DEPENDTS*ADMINSCR
 INTERC7= ESDG*BLACK
 INTERC8= ESDG*NUSEX
 INTERC9= ESDG*TERMENLT
 INTERC10= ESDG*SASVABAI
 INTERC11= ESDG*ADMINSCF
 INTERC12= ELACK*NUSEX
 INTERC13= ELACK*TERMENLT
 INTERC14= ELACK*SASVABAI
 INTERC15= ELACK*ADMINSCF
 INTERC16= NUSEX*TERMENLT
 INTERC17= NUSEX*SASVABAI
 INTERC18= NUSEX*ADMINSCF
 INTERC19= TERMENLT*SASVABAI
 INTERC20= TERMENLT*ADMINSCR
 INTERC21= SASVABAI*ADMINSCR;

/*
 //

APPENDIX B
DESCRIPTIVE ANALYSIS RESULTS

Frequency distributions and correlations used for descriptive analysis of the AD data set are contained in Tables XIV and XV.

The frequencies show that 92 percent of the AD data set were 17 to 21 years of age, 79 percent had a high school degree, 97 percent were single, and 98 percent were male. Even though BLACK and OTHER only represented 17 and 6 percent of the sample respectively, their criterion scores were significantly different compared to WHITE criterion scores. Thus, BLACK and OTHER emerged as predictors in some of the models. It is interesting to note that 40 percent of the sample achieved the paygrade E-5. Using achievement of E-5 rather than E-4 in the composite success criterion would produce greater variability on the criterion which may improve the models.

One third of the cases in the data did not score 190 or greater on the AD composite score. These cases are either people who were classified prior to correcting the ASVAB Forms 5, 6 and 7 misnaming problems, or people who migrated to the AD rating subsequent to service entry. This may partially explain the negative correlations these variables have with the criteria.

TABLE XIV
Selected Frequencies

DMDC RATE	FINAL RATING AS LISTED BY D.M.D.C.	PERCENT	CUM PERCENT
FREQUENCY	CUM FREQ		
AL	2820	100.000	100.000

SCREEN	FREQUENCY	SCREEN SCORE	PERCENT	CUM PERCENT
		CUM FREQ		
5	73	2	0.073	0.073
5	21	3	0.036	0.109
5	11	4	0.182	0.291
5	11	5	0.109	0.400
6	12	6	0.036	0.437
6	6	7	0.218	0.655
6	7	8	0.255	0.910
6	3	9	0.109	1.019
6	6	10	1.311	2.330
6	5	11	1.966	4.296
7	6	12	2.403	6.699
7	2	13	0.874	7.573
7	1	14	5.278	12.851
7	1	15	0.182	13.033
7	10	16	3.859	16.892
7	5	17	1.820	18.712
7	3	18	1.347	20.059
7	7	19	2.839	22.898
7	8	20	6.844	29.742
7	1	21	6.261	36.003
8	1	22	4.769	40.772
8	1	23	3.822	44.594
8	5	24	1.857	46.451
8	5	25	20.714	67.164
8	9	26	3.276	70.440
8	5	27	0.073	70.513
8	1	28	0.473	70.987
8	1	29	4.041	75.027
8	4	30	15.435	90.462
8	2	31	0.983	91.445
9	2	32	7.572	99.017
9	1	33	0.109	99.126
9	3	34	0.182	99.308
9	4	35	0.146	99.454
9	1	36	0.036	99.490
9	1	37	0.036	99.526
9	1	38	0.510	100.000

AFCT GFCUPS	(5, 4C, 4B, 4A, 3B, 3A, 2, 1)	PERCENT	CUM PERCENT
FREQUENCY	CUM FREQ		
1	4	0.142	0.142
2	6	2.163	2.305
3	28	9.929	12.234
4	59	21.241	33.475
5	79	28.191	61.667
6	54	19.326	80.993
7	50	17.908	98.901
8	31	1.099	100.000

ENTRYAGE	AGE OF INDIVIDUAL AT TIME OF ENTRY		PERCENT	
	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
17	388	388	13.759	13.759
18	122	510	43.440	57.199
19	594	2207	21.064	78.262
20	263	2470	9.326	87.589
21	136	2606	4.823	92.411
22	75	2681	2.660	95.071
23	49	2730	1.738	96.809
24	28	2758	0.993	97.801
25	22	2780	0.780	98.582
26	11	2791	0.390	98.972
27	12	2803	0.426	99.397
28	8	2811	0.284	99.681
29	8	2819	0.284	99.965
30	1	2820	0.035	100.000

ENTRYAGE	ENTRY PAY GRADE (E00--011)		PERCENT	
	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
1	2375	2375	84.220	84.220
2	276	2654	9.894	94.113
3	166	2820	5.887	100.000

TERMINIT	TERM OF ENLISTMENT (NO. OF YEARS)		PERCENT	
	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
3	1	1	0.035	0.035
4	1	2	0.035	0.071
5	269	2694	95.461	95.532
6	125	2820	4.433	100.000

SERVACCS	SERVICE OF ACCESSION (NAVY, 2)		PERCENT	
	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
2	2715	2715	96.277	96.277
8	105	2820	3.723	100.000

CEYEC	CONVERTED NUMBER OF YEARS OF EDUCATION		PERCENT	
	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
3.5	1	1	0.035	0.035
8	4	5	0.142	0.177
9	27	32	0.957	1.135
10	143	175	5.071	6.206
11	285	460	10.106	16.312
11.5	122	582	4.326	20.638
12	216	798	76.773	97.411
12.5	26	824	1.028	98.440
13	26	850	0.922	99.362
14	7	857	0.248	99.610
16	11	868	0.390	100.000

ESDG	HIGH-SCHOOL GRADUATE (1) V. OTHER (0)		PERCENT	
	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	582	582	20.638	20.638
1	2238	2820	79.362	100.000

DEPENDENTS	SINGLE, NC DEPENDENTS (0), OTHERWISE (1) FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	2738	2738	97.092	97.092
1	82	2820	2.908	100.000

NCSEX	(1) MALE, (0) FEMALE. FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	64	64	2.270	2.270
1	2756	2820	97.730	100.000

ENTRYGRP	ENTRY GROUP CLASSIFICATIONS FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
1	1166	1166	41.348	41.348
2	128	1294	4.539	45.887
3	1316	2610	46.667	92.553
7	210	2820	7.447	100.000

RACE	(1) WHITE, (2) BLACK, (3) OTHER FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
1	2184	2184	77.447	77.447
2	468	2652	16.596	94.043
3	168	2820	5.957	100.000

ALMINSCE	AD ASVAB COMECSITE SCREEN FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	945	945	33.511	33.511
1	1875	2820	66.489	100.000

RANIAL1	VAR TO ALLOW A RANDOM 50-50 SPLIT FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	1380	1380	48.936	48.936
1	1440	2820	51.064	100.000

ISC3	INTER-SERVICE SEPARATION CODE FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	1106	1106	39.220	39.220
1	1495	2601	53.014	92.234
60	22	2623	0.780	93.014
61	6	2629	0.213	93.227
63	1	2630	0.035	93.262
64	7	2637	0.248	93.511
65	61	2698	2.163	95.674
67	14	2712	0.496	96.170
71	7	2719	0.248	96.418
73	15	2734	0.532	96.950
74	1	2735	0.035	96.986
75	2	2737	0.071	97.057
76	7	2744	0.248	97.305
78	22	2766	0.780	98.085
80	4	2770	0.142	98.227
82	20	2790	0.709	98.936
86	30	2820	1.064	100.000

(1) INTERIM AFTER SEP 78, OTHERWISE (0)					
DATEENL1	FREQUENCY	CUM FREQ	PERCENT	CUM	PERCENT
0	2543	2543	90.177		90.177
1	277	2820	9.823		100.000

DMDC-BASFI HIGHEST PAY-GRADE ATTAINED					
PAYGRADE	FREQUENCY	CUM FREQ	PERCENT	CUM	PERCENT
1	110	110	3.901		3.901
2	108	218	3.830		7.730
3	231	449	8.191		15.922
4	1259	1708	44.645		60.567
5	1110	2818	39.362		99.929
6	2	2820	0.071		100.000

NHRC FILE--HIGHEST PAYGRADE ATTAINED					
NUHYPAY	FREQUENCY	CUM FREQ	PERCENT	CUM	PERCENT
1	17	17	0.603		0.603
2	95	112	3.369		3.972
3	220	332	7.801		11.773
4	1401	1733	49.681		61.454
5	1087	2820	38.546		100.000

SUCCESS ON LCS CRITERION (1)					
SUCCTAF	FREQUENCY	CUM FREQ	PERCENT	CUM	PERCENT
0	393	393	13.936		13.936
1	2427	2820	86.064		100.000

HIGH PAYGRADE SUCCESS CRITERION.					
SUCCPAYG	FREQUENCY	CUM FREQ	PERCENT	CUM	PERCENT
0	474	474	16.809		16.809
1	2346	2820	83.191		100.000

REENLISTMENT ELIGIBILITY CRITERION.					
SUCCREUF	FREQUENCY	CUM FREQ	PERCENT	CUM	PERCENT
0	293	293	10.390		10.390
1	2527	2820	89.610		100.000

SUCCESS ON COMEC SITE CRITERION (1)					
SUCCESS2	FREQUENCY	CUM FREQ	PERCENT	CUM	PERCENT
0	655	655	23.227		23.227
1	2165	2820	76.773		100.000

MONTHS OF ICTL. ACTIVE FED. MILIT. SERV.					
TAFMS1	FREQUENCY	CUM FREQ	PERCENT	CUM	PERCENT
2	1	1	0.035		0.035
3	4	5	0.142		0.177
4	2	7	0.071		0.248
5	6	13	0.213		0.461
6	1	14	0.035		0.496
7	1	15	0.106		0.603
8	1	16	0.284		0.887
9	1	17	0.319		1.206
10	1	18	0.177		1.383
11	1	19	0.177		1.560
12	1	20			
13	1	21			
14	1	22			
15	1	23			
16	1	24			
17	1	25			
18	1	26			
19	1	27			
20	1	28			
21	1	29			
22	1	30			
23	1	31			
24	1	32			
25	1	33			
26	1	34			
27	1	35			
28	1	36			
29	1	37			
30	1	38			
31	1	39			
32	1	40			
33	1	41			
34	1	42			
35	1	43			
36	1	44			
37	1	45			
38	1	46			
39	1	47			
40	1	48			
41	1	49			
42	1	50			
43	1	51			
44	1	52			
45	1	53			
46	1	54			
47	1	55			
48	1	56			
49	1	57			
50	1	58			
51	1	59			
52	1	60			
53	1	61			
54	1	62			
55	1	63			
56	1	64			
57	1	65			
58	1	66			
59	1	67			
60	1	68			
61	1	69			
62	1	70			
63	1	71			
64	1	72			
65	1	73			
66	1	74			
67	1	75			
68	1	76			
69	1	77			
70	1	78			
71	1	79			
72	1	80			
73	1	81			
74	1	82			
75	1	83			
76	1	84			
77	1	85			
78	1	86			
79	1	87			
80	1	88			
81	1	89			
82	1	90			
83	1	91			
84	1	92			
85	1	93			
86	1	94			
87	1	95			
88	1	96			
89	1	97			
90	1	98			
91	1	99			
92	1	100			
93	1	101			
94	1	102			
95	1	103			
96	1	104			
97	1	105			
98	1	106			
99	1	107			
100	1	108			
101	1	109			
102	1	110			
103	1	111			
104	1	112			
105	1	113			
106	1	114			
107	1	115			
108	1	116			
109	1	117			
110	1	118			
111	1	119			
112	1	120			
113	1	121			
114	1	122			
115	1	123			
116	1	124			
117	1	125			
118	1	126			
119	1	127			
120	1	128			
121	1	129			
122	1	130			
123	1	131			
124	1	132			
125	1	133			
126	1	134			
127	1	135			
128	1	136			
129	1	137			
130	1	138			
131	1	139			
132	1	140			
133	1	141			
134	1	142			
135	1	143			
136	1	144			
137	1	145			
138	1	146			
139	1	147			
140	1	148			
141	1	149			
142	1	150			
143	1	151			
144	1	152			
145	1	153			
146	1	154			
147	1	155			
148	1	156			
149	1	157			
150	1	158			
151	1	159			
152	1	160			
153	1	161			
154	1	162			
155	1	163			
156	1	164			
157	1	165			
158	1	166			
159	1	167			
160	1	168			
161	1	169			
162	1	170			
163	1	171			
164	1	172			
165	1	173			
166	1	174			
167	1	175			
168	1	176			
169	1	177			
170	1	178			
171	1	179			
172	1	180			
173	1	181			
174	1	182			
175	1	183			
176	1	184			
177	1	185			
178	1	186			
179	1	187			
180	1	188			
181	1	189			
182	1	190			
183	1	191			
184	1	192			
185	1	193			
186	1	194			
187	1	195			
188	1	196			
189	1	197			
190	1	198			
191	1	199			
192	1	200			
193	1	201			
194	1	202			
195	1	203			
196	1	204			
197	1	205			
198	1	206			
199	1	207			
200	1	208			
201	1	209			
202	1	210			
203	1	211			
204	1	212			
205	1	213			
206	1	214			
207	1	215			
208	1	216			
209	1	217			
210	1	218			
211	1	219			
212	1	220			
213	1	221			
214	1	222			
215	1	223			
216	1	224			

TABLE XV
Selected Correlations

	IAFMS1	SUCCTAF	SUCCESS2
AFQTGFES	-0.08930 0.0001	-0.06108 0.0012	-0.05137 0.0064
AFQTECNT	-0.07755 0.0001	-0.05346 0.0045	-0.04182 0.0264
ENTRYAGE	0.05518 0.0034	0.02593 0.1686	0.05698 0.0025
ENTFFAYG	0.03578 0.0575	-0.00926 0.6230	0.01083 0.5653
TEFMENIT	0.14720 0.0001	0.02116 0.2614	0.05189 0.0059
CHYEC	0.07522 0.0001	0.07554 0.0001	0.11471 0.0001
HSIG	0.09918 0.0001	0.12117 0.0001	0.15525 0.0001
NUSEX	0.08426 0.0001	0.04868 0.0097	0.01204 0.5229
WHITE	-0.10983 0.0001	-0.06035 0.0013	-0.04767 0.0114
BLACK	0.10220 0.0001	0.05290 0.0050	0.02641 0.1608
CIFER	0.03329 0.0771	0.02342 0.2139	0.04265 0.0235
SCREEN	-0.00478 0.8022	0.07461 0.0001	0.08891 0.0001
ALCCMECS	-0.05463 0.0037	-0.02132 0.2578	0.00440 0.8153
ADMINSCH	-0.07581 0.0001	-0.02971 0.1147	-0.02934 0.1192
SASVAEAD	0.00263 0.8888	0.01025 0.5864	0.01356 0.4717
SASVAEAI	-0.06941 0.0002	-0.03415 0.0698	-0.00654 0.7285
SASVAEAR	-0.05163 0.0061	-0.02568 0.1727	-0.01734 0.3574
SASVAEEI	-0.03140 0.0955	-0.01303 0.4893	0.01707 0.3648
SASVAEGI	-0.01535 0.4152	-0.02107 0.2633	-0.00708 0.7070
SASVAEMC	-0.06570 0.0005	-0.04088 0.0300	-0.02361 0.2101
SASVAEMK	-0.03166 0.0928	0.00698 0.7112	0.02288 0.2245

SASVAENC	-0.03869 0.0399	-0.01504 0.4246	-0.00541 0.7741
SASVAESI	-0.03834 0.0418	-0.00090 0.9618	-0.00387 0.8370
SASVAESF	-0.04680 0.0129	-0.02609 0.1660	-0.00735 0.6964
SASVAEGS	-0.03464 0.0659	-0.02632 0.1622	-0.01036 0.5822
SASVAEWK	-0.06134 0.0011	-0.05225 0.0055	-0.04783 0.0111
INTER01	0.05262 0.0052	0.03656 0.0523	0.04814 0.0106
INTER02	0.05869 0.0018	0.02943 0.1182	0.04022 0.0327
INTER03	0.07253 0.0001	0.03665 0.0516	0.04668 0.0132
INTER04	0.06297 0.0008	0.03358 0.0746	0.04590 0.0148
INTER05	0.06033 0.0013	0.03101 0.0996	0.04114 0.0289
INTER06	0.01791 0.3417	0.02303 0.2215	0.02299 0.2222
INTER07	0.09619 0.0001	0.05711 0.0024	0.03899 0.0384
INTER08	0.11832 0.0001	0.12665 0.0001	0.14906 0.0001
INTER09	0.12282 0.0001	0.11899 0.0001	0.16000 0.0001
INTER10	0.07658 0.0001	0.10700 0.0001	0.14937 0.0001
INTER11	0.00621 0.7416	0.04010 0.0332	0.06994 0.0002
INTER12	0.10322 0.0001	0.05193 0.0058	0.02540 0.1775
INTER13	0.10985 0.0001	0.05469 0.0037	0.02934 0.1153
INTER14	0.10585 0.0001	0.05537 0.0033	0.03218 0.0876
INTER15	0.06099 0.0012	0.03034 0.1073	0.00764 0.6852
INTER16	0.14975 0.0001	0.05095 0.0068	0.03845 0.0412
INTER17	-0.00842 0.6550	0.00025 0.9896	0.00180 0.9240
INTER18	-0.05198 0.0058	-0.01436 0.4460	-0.02373 0.2077

INTER19	0.00629 0.7384	-0.02271 0.2280	0.01637 0.3849
INTER20	-0.05623 0.0028	-0.02818 0.1347	-0.02189 0.2452
INTER21	-0.08291 0.0001	-0.03788 0.0443	-0.02863 0.1286

Note: The first number is the correlation between the predictor and the criterion, the second number is the significance level.

APPENDIX C

REGRESSION ANALYSIS PROGRAMS

Regression analysis attempts to predict or explain the values of the criterion variable with one or more predictor variables. The following sections expand upon the discussion of regression analysis presented in Chapter IV.

A. REQUIREMENTS AND ASSUMPTIONS

When conducting regression analysis, certain requirements must be met or assumed. One of these requirements is the use of quantitative variables.⁵ Application of regression procedures also requires normality (the value of the dependent variable must be normally distributed at each value of the independent variable), homoscedasticity (the variation around the regression line must be constant for all values of the independent variable), and independence of error (the residual difference between an observed and predicted value of the dependent variable must be independent for each value of the predictor variable). Another requirement of linear regression is that a straight-line or linear relationship exist between each independent variable and the dependent variable. For purposes of this study, and based on initial investigation, these requirements are assumed to be met. However, an extensive effort to evaluate these assumptions by transforming the variables or employing complex statistical analysis packages has not been conducted.

⁵The inclusion of qualitative or categorical variables in regression models may be accommodated through the use of dummy variables.

E. STEPWISE REGRESSION

The SAS Stepwise process considers each of the candidate independent variables for inclusion in the model by determining the contribution the variable makes to the model. This determination is accomplished by calculating the partial F statistic for the variable, and adding it to the model if it meets the specified entry significance level. After a variable is added, the stepwise method then looks at all the variables in the model and deletes any variable that does not provide an F statistic sufficient to meet the specified significance level for remaining in the model. This process of adding and deleting variables continues until none of the variables has an F statistic significant to enter or leave the model.⁶ [Ref. 12]

C. LINEAR REGRESSION

Simple linear regression is concerned with finding the statistical model or equation that best "fits" the original data. This is accomplished by defining a straight line that minimizes the differences between the actual value of the dependent variable and the value that would be predicted from the fitted line of regression. The SAS Regression procedure uses a mathematical technique, the least-squares method, to produce such an equation for the best linear model. This equation provides the intercept and slope of the sample predictor variable. With multiple linear regression, these slopes represent the unit change in the dependent variable per unit change in the independent variable, taking into account the effects of the other independent variables, and are referred to as net regression coefficients. The sample regression coefficients of the predictor

⁶This study used the SAS Stepwise default significance level of .15 for variables to enter or remain in the model.

variables are then used as estimates of the respective population parameters. For illustration, the program used to validate Model A is provided in Table XVI.

TABLE XVI
Sample Validation Program

```
//ADVALLE JCB (2807,C110),'D CSLUND, SMC 1763',CLASS=B
//*MAIN CFG=NPGVM1.28C7P
//EXEC SAS
//FILEIN ID DISP=SHR,DSN=MSS.S2807.ADALL4
//SYSIN ID *
CPTICNS IS=80 NOCENTIF;

* THIS PROGRAM CALCULATES THE VALIDITY OF A REGRESSION
  MODEL THROUGH THE USE OF CROSS-VALIDATION AND DOUBLE
  CROSS-VALIDATION TECHNIQUES.;

DATA DATA1;
  SET FILEIN.ADALL4;

* THE RANDOM VARIABLE CREATED IN 'ADNEWVAR' IS NOW USED
  TO SPLIT THE DATA APPROXIMATELY IN HALF. 'DERIVA' IS
  THE DERIVATION SAMPLE AND 'VALIDA' IS THE HOLD-OUT OR
  VALIDATION SAMPLE.;

DATA DERIVA;
  SET DATA1;
  IF RAN1AIL1 = 1;
DATA VALIDA;
  SET DATA1;
  IF RAN1AIL1 = 0;

* A FICK REGRESSION IS NOW RUN ON DERIVA TO COMPUTE AND
  OUTPUT THE PARAMETER ESTIMATES (BETAS) THAT RESULT
  FROM THE REGRESSION. THE BETAS ARE WRITTEN TO THE DATA-
  SET WORK.EETAD. THE MODEL IS GIVEN THE LABEL 'TAFMHATV';

PROC REG DATA=DERIVA CUTEST=BETAD;
  TAFMHATV:MODEL TAFMS1 = ADMINSCR TERMENTL DEPENDTS BLACK
                                HSDG      OTHER      NUSEX / STE;
TITLE REGRESSING ON DERIVA;

* THE NEXT STEP IS TO APPLY THE REGRESSION FORMULA (THE
  BETAS) TO THE DATA IN THE VALIDATION SAMPLE AND CALCULATE
  THE PREDICTED SCORE FOR EACH CASE IN VALIDA. THE PRED-
  ICTED SCORES ARE WRITTEN TO WORK.PREDTAFV. SAS USES THE
  MODEL LABEL (TAFMHATV) AS THE VARIABLE NAME FOR THE VALIDA
  PREDICTED SCORES. THE SCORE PROCEDURE PRODUCES NO
  PRINTED OUTPUT.;

PROC SCORE OUT=PREDTAFV TYPE=OIS SCORE=BETAD
  DATA=VALIDA PREDICT;
  VAR ADMINSCR TERMENTL DEPENDTS BLACK
      HSDG      OTHER      NUSEX;

* THE FIRST VALIDITY COEFFICIENT IS NOW CALCULATED BY FINI-
  SHING THE CORRELATION BETWEEN VALIDA'S ACTUAL SCORES AND
  VALIDA'S PREDICTED SCORES.;

PROC CORR DATA=PREDTAFV;
  VAR TAFMS1 TAFMHATV;
TITLE FIRST VALIDITY COEFFICIENT.;
```

* NOW TO REPEAT THE PROCESS TO UTILIZE THE DOUBLE CFCSS-
 VALIDATION TECHNIQUE. THIS TIME A REGRESSION IS RUN
 ON VALIDA AND THE RESULTING BETAS (BETAV) ARE USED TO
 PREDICT THE SCORES OF THE CASES IN DERIVA. DERIVA'S
 ACTUAL AND PREDICTED SCORES ARE THEN CORRELATED TO
 FIND THE SECOND VALIDITY COEFFICIENT.;

```
PROC REG DATA=VALIDA CTEST=BETAV;
  TAFMHATC:MODEL TAFMS1 = ADMINSCR TERMENLT DEPENDTS BLACK
                        HSDG      OTHER      NUSEX / STE;
TITLE REGRESSING ON VALIDA;
```

```
PROC SCORE COT=PREDTAFD TYPE=CIS SCORE=BETAV
  DATA=DERIVA PREDICT;
  VAR ADMINSCR TERMENLT DEPENDTS BLACK
      HSDG      OTHER      NUSEX;
```

```
PROC CORR DATA=PREDTAFD;
  VAR TAFMS1 TAFMEATD;
TITLE SECOND VALIDITY COEFFICIENT;
```

```
/*
//
```

APPENDIX D

DISCRIMINANT ANALYSIS PROGRAMS

Discriminant Analysis allows observations to be classified into two or more groups on the basis of one or more numeric variables. The following sections expand upon the discussion of discriminant analysis presented in Chapter IV. For illustration, Table XVII shows the program used to produce the classification matrices for the derivation and validation samples for Model A.

A. REQUIREMENTS AND ASSUMPTIONS

As was the case with regression analysis, discriminant analysis also requires that certain basic assumptions be met. First, the observations in the data set should be members of two or more mutually exclusive groups. Therefore, the groups must be defined so that each case will belong to only one group. Another statistical property required of discriminating variables is that they may not be linear combinations of other variables. Thus, the sum or average of several variables may not be used along with those variables. There are three other assumptions to be considered. The population covariance matrices must be equal for each group, each group is to be drawn from a population which has a multivariate normal distribution, and discriminating variables must be measured at the interval or ratio levels. Ideally, these variables will be continuous, but they need not be. [Ref. 17] This study assumes these requirements have been met. However, an effort to evaluate these properties was not conducted since, in practice, the discriminant analysis technique is rather robust and can tolerate some deviation from these assumptions [Ref. 18].

E. DISCRIMINANT ANALYSIS

The first step of discriminant analysis is to weight and linearly combine the discriminating variables so that the groups will be as statistically distinct as possible. The derived equations, called discriminant functions, combine the group characteristics using a measure of generalized squared distance⁷ that will allow one to identify the group to which a case belongs or most closely resembles.

The classification process may assume that membership in a group has equal likelihood of occurring. However, it may be more desirable to incorporate the prior probability of group membership into the classification function to improve prediction accuracy or minimize the cost of prediction errors. In this study, membership in a success group was on the order of 80 percent. Therefore, it was appropriate to consider prior probabilities so that those cases predicted as unsuccessful would be classified as such only if strong evidence exists that they belong there.

The ultimate concern in developing a mathematical model is that it predict well or provide a reasonable description of the real world. Once a model is developed which provides satisfactory discrimination for cases of group membership, classification functions may be derived and applied to the classification of new cases with unknown group membership. A good test of the adequacy and accuracy of the discriminant model is the percentage of correct classifications, commonly called the "hit-rate". This test is accomplished by applying the classification function to the known cases from which the function was derived. The percentage of correctly

⁷The procedure conducted a likelihood ratio test of homogeneity of the within-group covariance matrices for each model. This test was statistically significant for each model. Therefore, the within-group matrices were used as the basis of the measure of generalized squared distance in developing the classification criterion. [Ref. 12]

classified cases provides an indication of the accuracy of the procedure and indirectly confirms the degree of group separation. The results may be depicted in a classification matrix.

When the sample size is large enough, as it is in this study, a further check of the classification accuracy may be conducted by randomly splitting the sample into two subsets. The classification function is derived on one subset and validated on the other subset. A comparison of the two hit-rates provides the measure of accuracy of the model. [Ref. 17]

TABLE XVII

Sample Discriminant Analysis Program

```
//DISCEGMS JOB (2807,C110),'D CSLUND, SMC 1763',CLASS=B
//*MAIN CRG=NP GVM1.28C7P
//EXEC SAS
//FILEIN ID DISP=SHR,DSN=MSS.S2807.ADALL4
//SYSIN ID *
CPTICNS IS=80 NOCENTEF:
```

* THIS PURPOSE OF THIS PROGRAM IS TO ALLOW THE VALIDITY OF A DISCRIMINANT MODEL TO BE INVESTIGATED. A CLASSIFICATION FUNCTION IS DERIVED FROM THE DERIVA SAMPLE AND THIS FUNCTION IS USED TO CLASSIFY THE CASES IN THE VALIDATION (OR HCLD-OUT) SAMPLE. THE TWO CLASSIFICATION MATRICES ARE THEN USED TO AILCW THE 'HIT RATE' ON EACH SAMPLE TO BE CALCULATED.;

```
DATA IATA1;
  SET FILEIN.ADALL4;
```

* USING THE RANDOM VARIABLE TO SPLIT THE SAMPLE APPROXIMATELY IN HALF.;

```
DATA DERIVA;
  SET IATA1;
  IF RANLATA1=1;
DATA VALIDA;
  SET IATA1;
  IF RANLATA1=0;
```

* CALCULATING THE CLASSIFICATION MATRIX FOR DERIVA AND WRITING OUT THE CLASSIFICATION FUNCTION DERIVED FROM DERIVA TO WORK.D.;

```
PROC DISCRIM DATA=DERIVA OUT=L PCCL=TEST;
  CLASS SUCCTAF;
  VAR DEPENDTS HSDG BLACK TERMENTLT
      NUSEX CTHEF ADMINSCR;
  PRICES PROPORTIONAL;
  TITLE DISCRIM ON DERIVA.;
```

* NOW THE CLASSIFICATION FUNCTION FROM DERIVA IS USED TO CLASSIFY THE CASES IN VALIDA.;

```
PROC DISCRIM DATA=D TESTDATA=VALIDA;
  TESTCLASS SUCCTAF;
  TITLE DERIVA'S FUNCTION APPLIED TO VALIDA.;
```

```
/*
//
```


APPENDIX E

UTILITY ANALYSIS PROGRAMS

This appendix provides further details of the information contained in Chapter V, and gives examples of the SAS programs and outputs.

A. CALCULATION OF CELL PROBABILITIES

The method used to calculate cell probabilities in this study depends on whether a regression or a discriminant model is being evaluated. A regression model can be viewed simply as a formula for calculating predicted scores, whereas a discriminant model actually classifies cases as predicted successes or predicted failures. Because of this difference, the calculation of cell probabilities is more complicated for regression models than for discriminant models.

1. Regression Models

A regression model and the data from which it was developed provide information on the predicted and actual scores for each case. In order to classify these cases into the four selection outcomes, the cut score on the predictor and the score on the criterion above which people are considered to be successful must be known. If the criterion is constructed as a dichotomous (success/fail) variable, then the cases assigned a value of "one" are considered successful and those with a value of "zero" are considered unsuccessful. If the criterion is a continuous variable (such as length of service) then a value on the scale must be chosen as the dividing line between success and failure.

The choice of the cut score is not such a simple matter, and cannot be arbitrarily assigned as can the distinction between success and fail. The choice of the cut score, as mentioned before, often depends on the desired selection ratio. In the absence of information on the desired selection ratio, cell probabilities are calculated for each of many possible cut scores, and a cut score is eventually chosen based on which set of cell probabilities maximizes the utility of the model. In a data set containing actual and predicted scores, different sets of cell probabilities can be calculated if each predicted score is considered to be a cut score. Table XVIII contains five pairs of actual and predicted scores which will be used to illustrate the method.

TABLE XVIII
Illustrative Actual and Predicted Scores

Actual Criterion Score	Predicted Criterion Score
50	44
44	46
49	47
46	49
49	50

In this illustration, cases who serve 48 months or longer are considered to be successful. Each different predicted score will be considered as a cut score and cell counts for each cut score will be calculated. If the cut score is 44 months, then all cases with a predicted score of 44 months or more will be accepted, and those with a predicted score of less than 44 months will be rejected. In this example, for a cut score of 44, all cases will be accepted. No one is rejected, therefore, valid negatives and false negatives will be zero. Of the five cases

AD-A152 127

THE DEVELOPMENT OF AN ENLISTMENT STANDARDS MODEL FOR
THE NAVY AVIATION MACHINIST'S MATE (AD) RATING(U) NAVAL
POSTGRADUATE SCHOOL MONTEREY CA D A OSLUND ET AL.

3/2

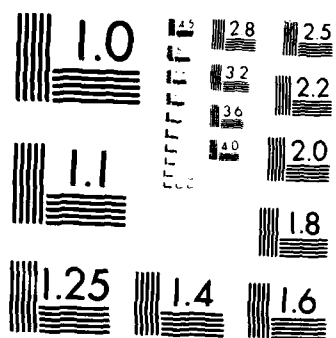
UNCLASSIFIED

JUN 84

F/G 5/9

NL

							END						
							FILED						
							FILE						



MICROCOPY RESOLUTION TEST CHART
 NATIONAL BUREAU OF STANDARDS-1963-A

accepted, three have actual LOS of 48 months or more (successes). Therefore, the number of valid positives is three. Two of the five cases accepted had actual LOS of less than 48 months (failures). Therefore, false positives will be two. Thus the first set of cell probabilities that result when the cut score is 44 are: $PVP = 3/5$, $PPF = 2/5$, $PFN = 0$ and $PVN = 0$. The next set of cell probabilities will result when 46 months is considered to be the cut score. One case had a predicted LOS of less than 46. Therefore, he would be rejected. His actual LOS is 50 months, so he was falsely rejected, i.e. $FN = 1$. No one else was rejected so $VN = 0$. Four cases had a predicted LOS of 46 or greater so all four would be accepted. Of these four, two had actual LOS of less than 48 months (FF), and two had actual LOS of 48 months or more (VP). Thus for a cut score of 46, $PVP = 2/5$, $PPF = 2/5$, $PFN = 1/5$ and $PVN = 0$. This process is repeated until five sets of cell probabilities (one for each different predicted score) are calculated.

2. Discriminant Models

In a discriminant model the criterion is a categorical (0,1) variable. The output from the SAS Discriminant procedure is a two by two table where the cases are predicted to be either a "zero" or a "one", and the prediction is compared to the actual score. Table XIX gives an abbreviated example of the output from the discriminant procedure.

The columns are the model's predicted scores for the 750 cases in this hypothetical sample. Here the model predicts that 300 of the cases will score "zero" on the criterion, and it predicts that 450 of the cases will score a "one" on the criterion. The rows are the actual scores of the cases. 250 people actually scored "zero" (failures) and

TABLE XIX
Illustrative Discriminant Example

		Predicted		
		0	1	Total
Actual	0	100	150	250
	1	200	300	500
Total		300	450	750

500 people actually scored "one" (successes). Because, in effect, the discriminant procedure chooses its own cut score, the four cell probabilities can be derived directly from the output. The predicted "ones" are people that the model classifies as accept. Of these 450, 150 actually failed so they are false positives, and the remaining 300 were successful, so they are valid positives. Of the 300 cases that the model would have rejected (predicted "zeros"), 100 were failures (valid negatives) and 200 were successes (false negatives). Again the cell probabilities are found by dividing each count by the number of cases. Therefore, $FVP = 300/750$, $FPF = 150/750$, $PFN = 200/750$ and $FVN = 100/750$. For a discriminant model, there is only one set of cell probabilities to be calculated.

E. ESTIMATION OF CELL UTILITIES

In order to calculate the overall utility of a model, utilities associated with each selection outcome need to be estimated. "Although the assignment of utility values to outcomes may very well be the 'Achilles Heel' of decision theory, it is not a problem that can be ignored by any institution that makes personnel decisions." [Ref. 19]

Ideally each selection outcome should have a uniquely estimated utility. Because of the difficulty in estimating

utilities for each outcome (particularly for the false and valid negatives), relative utilities are estimated. It is apparent that a person who is correctly selected (valid positive) has a positive worth to the organization. A reasonable estimate of this worth is the marginal product of the employee. In this study it is assumed that the navy compensates sailors at the full value of their marginal product, and the Billet Cost Model provides an estimate of the cost to the Navy of staffing a billet [Ref. 16]. Because relative utilities are the issue at this time, the utility of a valid positive (U1) is assigned the value of +1.

It is a reasonable assumption that the utility of a false positive is a negative number. As the employee was not judged to be successful, his marginal product was probably less than the marginal cost to keep him in the job. In addition a poor performer may adversely affect the performance and productivity of his fellow employees, and when he leaves, additional expense is necessary to find a replacement. On the other hand, it is unlikely that a poor performer does not contribute anything to the organization, and thus it is obviously difficult to estimate the magnitude of the disutility of a false positive. In this study a minor form of sensitivity analysis is undertaken to circumvent this estimation difficulty, and expected overall utilities are calculated for three different relative values of false positive utility (U2). These values are -.5, -1, and a relatively extreme assumption, -2.

The disutility of a false negative is also difficult to estimate, partly because it is not known what happens to the applicant after he is rejected. If the Navy rejects an applicant to the AD rating but accepts him in another rating where he is subsequently successful, then his disutility could be reasonably argued to be zero. If, however, the

Navy rejects him altogether when he would have been successful if selected, then the costs of attracting and testing him are wasted and additional costs are required to attract and test another applicant. These costs will depend on the state of the recruiting market at the time. If there are many good quality applicants then the disutility of rejecting a potentially successful applicant may be small. Again, as a type of sensitivity analysis, three relative values for the utility of a false negative (U_3) are considered: 0, -.25 and -.5.

It is not obvious that any utility should be assigned to U_4 , the utility of a valid negative. The person would have failed anyway, so nothing was gained by rejecting him. However, when viewed from an economist's viewpoint in relation to opportunity costs, the fact that the person was correctly rejected means that the organization did not have to bear the cost of incorrectly accepting someone who turns out to be unsuccessful. Thus, correctly rejecting an applicant is of equal and opposite utility to incorrectly accepting him. Therefore, $U_4 = -U_2$.

The use of relative utilities is a convention to simplify the estimation of cell utilities. In the above discussion relative utilities are estimated on the basis that the utility of a valid positive is +1. However, the values of U_1 through U_4 that are used in the formula for overall expected utility, (Equation 5.1), need to be expressed in actual dollars. As mentioned above, the Billet Cost Model is used to estimate the utility of a valid positive. The standard manyear cost of an E-4 Aviation Machinist's Mate is \$24,163. This cost comes from financial year 1983 data and represents the total cost to the Navy of creating and filling a job slot over one full year. [Ref. 16] A utility of +1 is therefore equivalent to +\$24,163, a utility of -.5 will be -\$12,082, and so on.

C. PROGRAMS USED TO CALCULATE UTILITIES

As mentioned in Section A. above, the calculation of cell probabilities for a regression model is a fairly tedious and repetitive procedure. This section contains three sample programs used to calculate the expected utility of a model. Explanatory comments are provided following each set of SAS statements. The first program (Table XX) computes the predicted criterion score for each case and writes the results out to a file called "RTYHATA". Table XXI shows part of the output from the first program. The second program's main purpose (Table XXII) is to calculate the cell probabilities that would result if each different predicted score were used as a cut score. The cell probabilities are written out to a file called "RTUTIIA". The program also calculates the expected utilities for one set of cell utilities and outputs the 30 largest utilities that result (Table XXIII). The third program (Table XXIV) calculates the utilities for six different sets of cell utilities.

As explained before, only one set of probabilities results from a discriminant model and these can be readily gained from the discriminant output. No programs were used to calculate the expected utilities of a discriminant model and these calculations were done by hand.

D. CALCULATION OF BASE LINE UTILITIES

As described in Chapter V, the utility of the Navy's original selection strategy (the base line utility) needs to be calculated in order for comparisons to be made. Observation 4 in Table XXIII demonstrates that when all the cases are accepted (41.0774 is the lowest predicted score), the selection ratio is obviously 1 and $PVP = .860638$ (which is the base rate) and $PPF = 1 - PVP = .139362$. No one is

rejected, therefore PFN and FVN are zero. The expected utility under these circumstances is:

$$EU = .860638(\$24,163) + .139362(-\$12,082) + 0 + 0 = \$19,112$$

As Table XXIII shows, the maximum utility occurs when the cut score is slightly higher than the lowest predicted score (there are five cases with a predicted score of less than 43.2692 in Table XXI). This maximum utility (\$19,135) is .12 percent greater than the base line utility of \$19,112.

TABLE XX
First Utility Analysis Program

```
//SEIUT111 JOB (2840,C104), 'SEI CLARK, SMC 1560', CLASS=E
//*MAIN CRG=NPGVM1.2840P
//EXEC SAS
//FILEIN DD DISP=SHR,DSN=MSS.S2807.ADALL4
//FILEOUT DD UNIT=3330V,MSVGP=PUB4A,DISP=(NEW,CATLG,DELETE),
//          DSN=MSS.S2840.RTYHATA,
//          LCB=(BLKSIZE=6400)
//SYSIN DD *
CPTICNS IS=80 NOCENTRE;

* THE PURPOSE OF THIS PROGRAM IS TO CALCULATE THE PREDICTED
  SCORE FOR EACH CASE (USING THE MODEL DEVELOPED PRE-
  ICLUSIV), AND TO WRITE OUT THE ACTUAL AND PREDICTED
  SCORES TO A FILE IN MASS STORAGE.;

DATA DATA1;
  SET FILEIN.ADALL4;
  RENAME TAFMS1=Y;

* RENAMING THE CRITERION VARIABLE;

PROC REG DATA=DATA1 CUIEST=BETAS;
  YHAT=MODEL Y =
  DEFENDIS HSDG ELACK OTHER NUSEX TERMENLT ADMINSCR / STE;
  TITLE ELCK REGRESSION TO OUTPUT BETAS.;

PROC SCORE OUT=PREDY TYPE=OLS SCORE=BETAS DATA=DATA1 PREDICT;
  VAR DEFENDIS HSDG ELACK OTHER NUSEX TERMENLT ADMINSCR;

* CALCULATES THE PREDICTED SCORES, AND WRITES THEM TO
  DATASET 'PREDY'.
  NCIE: THE SCORE PROCEDURE TAKES THE MODEL LABEL (YHAT)
        AND USES THAT LABEL AS THE VARIABLE NAME FOR THE
        PREDICTED SCORE.;

DATA PREDY2;
  SET PREDY;
  KEEP YHAT Y SUCCTAF;

PROC SORT DATA=PREDY2 OUT=FILEOUT.RTYHATA;
  BY YHAT;

* SORTS THE OUTPUT FILE INTC ASCENDING YHAT ORDER,
  AND WRITES OUT THE SORTED DATA TO MASS STORAGE.;

DATA TEST;
  SET FILEOUT.RTYHATA;
  IF _N_ LE 10 OR (_N_ GT 1270 AND _N_ LE 1280)
  OR _N_ GT 2790;

PROC PRINT DATA=TEST SPLIT=*;
  LABEL Y=ACTUAL*CRITERION*SCORE
        YHAT=PREDICTED*CRITERION*SCORE
        SUCCTAF=SUCCESS CN*CRITERION;
  TITLE THE FIRST 10, MIDDLE 10 AND LAST 30 OBS OF RTYHATA;
  TITLE2;
  TITLE3: NCIE: SORTED IN ASCENDING ORDER OF YHAT.;

PROC UNIVARIATE DATA=FILEOUT.RTYHATA PLOT;
  VAR YHAT Y SUCCTAF;
  TITLE STATS OF THE ACTUAL AND PREDICTED CRITERION SCORES;
```

TABLE XXI

Partial Output from the First Utility Program

THE FIRST 10, MIDDLE 10 AND LAST 30 OBS OF RIYHATA

NOTE: SORTED IN ASCENDING ORDER OF YHAT.

CES	ACTUAL CRITERION SCORE	SUCCESS ON CRITERION	PREDICTED CRITERION SCORE
1	4.5	0	41.0774
2	2.5	0	41.0774
3	6.6	1	42.0297
4	1.2	0	42.0297
5	1.6	0	42.8960
6	2.1	0	43.2692
7	1.1	0	43.2692
8	5.4	1	43.2692
9	4.8	1	43.2692
10	3.5	1	43.2692
11	3.5	1	48.7601
12	5.1	1	48.7601
13	4.7	1	48.7601
14	4.8	1	48.7601
15	6.5	1	48.7601
16	7.1	1	48.7601
17	4.3	1	48.7601
18	3.3	0	48.7601
19	4.4	1	48.7601
20	4.9	1	48.7601
21	5.1	1	56.8469
22	5.5	1	56.8469
23	5.5	1	56.8469
24	5.5	1	56.8469
25	5.5	1	56.8469
26	7.0	1	56.8469
27	5.1	1	56.8469
28	5.0	1	56.8469
29	4.6	1	56.8469
30	5.0	1	56.8469
31	6.0	1	57.7992
32	5.4	1	57.7992
33	5.7	0	57.7992
34	5.5	1	57.7992
35	5.5	1	57.7992
36	5.5	1	57.7992
37	6.0	1	57.7992
38	6.4	1	57.7992
39	6.4	1	57.7992
40	6.4	1	57.7992
41	5.5	1	57.7992
42	5.5	1	57.7992
43	6.4	1	57.7992
44	6.4	1	57.7992
45	6.4	1	57.7992
46	6.4	1	57.7992
47	6.4	1	57.7992
48	6.4	1	57.7992
49	6.4	1	57.7992
50	6.4	1	57.7992
51	6.4	1	57.7992
52	6.4	1	57.7992
53	6.4	1	57.7992
54	6.4	1	57.7992
55	6.4	1	57.7992
56	6.4	1	57.7992
57	6.4	1	57.7992
58	6.4	1	57.7992
59	6.4	1	57.7992
60	6.4	1	57.7992
61	6.4	1	57.7992
62	6.4	1	57.7992
63	6.4	1	57.7992
64	6.4	1	57.7992
65	6.4	1	57.7992
66	6.4	1	57.7992
67	6.4	1	57.7992
68	6.4	1	57.7992
69	6.4	1	57.7992
70	6.4	1	57.7992
71	6.4	1	57.7992
72	6.4	1	57.7992
73	6.4	1	57.7992
74	6.4	1	57.7992
75	6.4	1	57.7992
76	6.4	1	57.7992
77	6.4	1	57.7992
78	6.4	1	57.7992
79	6.4	1	57.7992
80	6.4	1	57.7992
81	6.4	1	57.7992
82	6.4	1	57.7992
83	6.4	1	57.7992
84	6.4	1	57.7992
85	6.4	1	57.7992
86	6.4	1	57.7992
87	6.4	1	57.7992
88	6.4	1	57.7992
89	6.4	1	57.7992
90	6.4	1	57.7992
91	6.4	1	57.7992
92	6.4	1	57.7992
93	6.4	1	57.7992
94	6.4	1	57.7992
95	6.4	1	57.7992
96	6.4	1	57.7992
97	6.4	1	57.7992
98	6.4	1	57.7992
99	6.4	1	57.7992
100	6.4	1	57.7992

TABLE XXII
Second Utility Analysis Program

```
//SEIUTII2 JOB (2840,C104), 'SEI CLARK, SMC 1560', CLASS=E
//*MAIN CFG=NPGVM1.2840P
// EXEC SAS
//SAS.LCFCR LD SPACE=(CYL,(12,4))
//FILEIN DD DISP=SHR,DSN=MSS.S2840.RTYHATA
//FILEOUT DD UNIT=3330V,MSVGF=SUB4A,DISP=(NEW,CATLG,DELETE),
//          DSN=MSS.S284C.RTUTIIA,
//          LCB=(BLKSIZE=6400)
//SYSIN DD *
CPTICNS IS=80 NOCENTIF;
```

* THE PURPOSE OF THIS PROGRAM IS TO WRITE OUT A FILE TO MASS STORAGE WHICH CONTAINS THE VALUES OF PVP, PFP, PFN AND PVN THAT RESULT WHEN EACH PREDICTED SCORE IS USED TO SEPARATE THE CASES INTO ACCEPT AND REJECT GROUPS (IE. OUTPUT THE CELL EFFICIENCIES THAT RESULT WHEN EACH PREDICTED SCORE IS USED AS A CUTTING SCORE).

THE INPUT FILE CONTAINS 3 VARIABLES, AND THE OBSERVATIONS (OR CASES) ARE SORTED IN ASCENDING ORDER OF 'YHAT'. 'YHAT' IS THE PREDICTED LOS (FROM THE MODEL DEVELOPED EARLIER) OF EACH CASE, 'Y' IS THE ACTUAL LOS IN MONTHS AND 'SUCC' IS A DUMMY VARIABLE WHERE EACH CASE IS CATEGORIZED AS A SUCCESS (1) OR AS A FAILURE (0) .;

```
DATA DATA1;
  SET FILEIN.RTYHATA;
  DECF Y;
  RENAME SUCC = Y;
```

* THE DATA IS READ IN AND THE ACTUAL LOS IN MONTHS VARIABLE IS RECEIVED AND THE DUMMY VARIABLE IS RENAMED 'Y' .;

```
PROC SUMMARY DATA=DATA1;
  VAR Y;
  OUTPUT OUT=DATA2 SUM=NSUCC N=NCASE;
```

* HERE THE NUMBER OF SUCCESSFUL CASES IN THE DATA (NSUCC) IS FOUND BY SUMMING THE 1'S AND 0'S IN VARIABLE 'Y'. ANOTHER VARIABLE 'NCASE' IS CREATED WHICH IS THE NUMBER OF CASES IN THE DATA. THESE TWO VARIABLES (EACH A SINGLE NUMBER) ARE WRITTEN TO DATA SET WORK.DATA2.;

```
DATA DATA3;
  IF N EQ 1 THEN SET DATA2;
  NFALL = NCASE - NSUCC;
  SET DATA1;
```

* THE VARIABLES NCASE, NSUCC AND NFALL (THE NUMBER OF UNSUCCESSFUL CASES IN THE DATA) ARE ADDED TO DATA1. NCASE, NSUCC AND NFALL ARE EACH SINGLE NUMBERS THAT ARE REPEATED FOR EACH OBSERVATION. EG. NCASE IS A COLUMN OF 500'S (SAY), NSUCC IS A COLUMN OF 325'S (SAY) AND THEREFORE NFALL IS A COLUMN OF 175'S.;

```

DATA DATA4:
SET DATA3:
U1= 24163; U2= -12082; U3= -6041; U4= 12082;
RETAIN NZERO 0;
RETAIN LASTYHAT 0;
IF YHAT NE LASTYHAT THEN LINK CALCS; ELSE LINK ZEFCS;
IF Y=C THEN NZERC=NZERO+1;
LASTYHAT=YHAT;
RETURN;
CALCS: VP = NSUCC-(N-1-NZERO);
FP = NFAIL-NZERO;
FN = N-1-NZERO;
VN = NZERC;
UTIL = (U1*VP + U2*FP + U3*FN + U4*VN)/NCASE;
SRATIO = (VF+FP)/NCASE;
SUCCRATE = VF/(VP+FP);
RETURN;
ZEFCS: VF = 0; FP = 0; FN = 0; VN = C;
UTIL = 0; SRATIO = 0; SUCCRATE = 0;
RETURN;

```

* THIS IS THE HEART OF THE PROGRAM WHERE SUBTLE LOGIC IS EMPLOYED. 'NZERO' IS A COUNTER WHICH COUNTS THE NUMBER OF 0'S IN THE 'Y' VARIABLE DOWN TO AND INCLUDING THE LINE (OBSERVATION) CONTAINING THE 'CURRENT' CUTTING SCORE. FOR EXAMPLE, IF THERE ARE 150 ZEROS AND 250 ONES AMONG THE FIRST 400 OBS. OF 'Y', THEN THE 400TH OBS. OF 'NZERC' WILL BE 150. IF THE 401ST OBS. OF 'Y' IS A ZERO THEN THE 401ST OBS. OF 'NZERO' WILL BE 151. TO CONTINUE THE EXAMPLE, BECAUSE THE INPUT DATA IS SORTED IN ASCENDING ORDER OF 'YHAT', THE 400 CASES PRECEDING THE 401ST CASE (WHICH IS THE CURRENT CUTTING SCORE), WOULD ALL BE CLASSIFIED AS REJECT BECAUSE THEIR PREDICTED SCORE IS LESS THAN THE CUTTING SCORE. THE 400TH OBS. OF 'NZERO' TELLS US HOW MANY OF THESE REJECTED CASES WERE FAILURES AND THEREFORE VN = NZERC. 'NFAIL' IS THE TOTAL NUMBER OF CASES THAT FAILED, THEREFORE NFAIL-VN (SAME AS NFAIL-NZERO) = FP. THE NUMBER OF CASES IN THE REJECTED 400 CASES (FN) IS THE CURRENT CES. (401) MINUS 1, MINUS THE NUMBER OF ZEROS, OR FN = 401-1-150 = 250. FINALLY, 'NSUCC' IS THE TOTAL NUMBER OF SUCCESSES, THEREFORE NSUCC-FN IS THE VALUE OF VP.

'LASTYHAT' IS USED TO PRECLUDE ANY ERRORS THAT WOULD BE GENERATED WHEN TWO OR MORE VALUES OF 'YHAT' ARE IDENTICAL. IF THE NEXT POTENTIAL CUTTING SCORE IS THE SAME AS THE PREVIOUS ONE, THEN NO CELL PROBABILITIES, ETC ARE CALCULATED, AND ZEROS ARE ASSIGNED.
NOTE: DUE TO THE USE OF THE KEYWORD 'RETAIN', THE VALUES OF NZERC AND LASTYHAT USED IN THE CALCULATIONS AND IN THE FIRST 'IF' STATEMENT ARE THE VALUES FROM THE PREVIOUS OBSERVATION.

THE DATA STEP ALSO INITIALIZES A SET OF INDIVIDUAL CELL UTILITIES (U1 - U4) AND CALCULATES THE OVERALL UTILITY ASSOCIATED WITH EACH CUTTING SCORE. ALSO THE SELECTION RATIO AND THE SUCCESS RATE RESULTING FROM EACH CUTTING SCORE ARE CALCULATED.

```

DATA DATA5:
SET DATA4:
PVP = VP/NCASE; PFP = FP/NCASE;
PFN = FN/NCASE; PVN = VN/NCASE;
KEEP YHAT UTIL PVP PFP PFN PVN SRATIO SUCCRATE;
RENAME YHAT = CSCCFE;
LABEL
CSCCFE=CUT SCORE ON PREDICTOR;

```

* CONVERTING THE CELL COUNTS TO PROBABILITIES.;

```
PROC SORT DATA=DATA5 OUT=FILECUT.HTUTILA;  
BY DESCENDING UTIL;
```

```
* SORTING BY UTIL BEFORE WRITING OUT THE PREVIOUSLY KEPT  
VARIABLES TO A FILE IN MASS STORAGE.;
```

```
DATA DATA6;  
SET FILECUT.HTUTILA;  
IF _N_ LE 30;
```

```
PROC PRINT DATA=DATA6;  
TITLE THE 30 LARGEST UTILITIES IN THE FILECUT.;  
TITLE2;  
TITLE3 THE BASE UTILITY IS 19112, AND THE;  
TITLE4;  
TITLE5 BASE LINE SUCCESS RATE IS 0.8606;
```

```
PROC PLOT DATA=DATA6;  
PLOT UTIL * CSCORE = '+' / VREF =19112;  
TITLE THE TCP 30 UTILITIES PLOTTED AGAINST CUTTING SCORE.;  
TITLE2;  
TITLE3 NOTE: THE HORIZ. LINE IS THE BASE LINE UTILITY,;  
TITLE4;  
TITLE5 IE. THE UTILITY RESULTING FROM THE NAVY'S;  
TITLE6;  
TITLE7 ORIGINAL SELECTION STRATEGY. (19112);
```

```
PROC PLOT DATA=DATA6;  
PLOT UTIL * SRATIC = '+' / VREF =19112;  
TITLE THE TCP 30 UTILITIES PLOTTED AGAINST SELECTION RATIC.;  
TITLE2;  
TITLE3 NOTE: THE HORIZ. LINE IS THE BASE LINE UTILITY,;  
TITLE4;  
TITLE5 IE. THE UTILITY RESULTING FROM THE NAVY'S;  
TITLE6;  
TITLE7 ORIGINAL SELECTION STRATEGY. (19112);
```

```
PROC PLOT DATA=FILECUT.HTUTILA;  
PLOT UTIL * SRATIC = '+' / VREF =19112;  
TITLE PLOTTING ALL UTILITIES AGAINST SELECTION RATIC.;  
TITLE2;  
TITLE3 NOTE: THE HORIZ. LINE IS THE BASE LINE UTILITY,;  
TITLE4;  
TITLE5 IE. THE UTILITY RESULTING FROM THE NAVY'S;  
TITLE6;  
TITLE7 ORIGINAL SELECTION STRATEGY. (19112);
```

```
/*  
//
```

TABLE XXIII

Partial Output from the Second Utility Program

THE 30 LARGEST UTILITIES IN THE FILEOUT.

THE BASE UTILITY IS 19112, AND THE

BASE LINE SUCCESS RATE IS 0.8606.

CES	CSCCRE	UTIL	SRATIO	SUCCRATE	PVP	PFP	PFN	PVN
1	4	191135	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
2	4	191129	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
3	4	191127	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
4	4	191111	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
5	4	189999	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
6	4	189998	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
7	4	189997	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
8	4	189996	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
9	4	189995	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
10	4	189994	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
11	4	189993	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
12	4	189992	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
13	4	189991	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
14	4	189990	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
15	4	189989	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
16	4	189988	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
17	4	189987	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
18	4	189986	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
19	4	189985	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
20	4	189984	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
21	4	189983	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
22	4	189982	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
23	4	189981	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
24	4	189980	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
25	4	189979	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
26	4	189978	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
27	4	189977	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
28	4	189976	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
29	4	189975	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001
30	4	189974	0.8606	0.8606	0.8606	0.1388	0.0001	0.0001

TABLE XXIV
Third Utility Analysis Program

```
//SEIUTII3 JOB (2840,C104), 'SEI CLARK, SMC 1560', CLASS=E
//*MAIN CRG=NPGVM1.2840P
// EXEC SAS
//FILEIN DD DISP=SHR,DSN=MSS.S2840.RTUTILA
//SYSIN DD *
CPTICNS IS=80 NOCENTEE;

* THIS PROGRAM EXPLCES THE EFFECTS OF USING DIFFERENT CELL
  UTILITIES FOR THE CALCULATION OF OVERALL UTILITY.;

DATA DATA1;
  SET FILEIN.RTUTILA;

  U1A= 1 : U2A= -.5 : U3A= 0 : U4A= .5 :
  U1E= 1 : U2E= -1 : U3E= 0 : U4E= 1 :
  U1C= 1 : U2C= -2 : U3C= 0 : U4C= 2 :
  U1D= 1 : U2D= -.5 : U3D= -.5 : U4D= .5 :
  U1F= 1 : U2F= -1 : U3F= -.5 : U4F= 1 :
  U1I= 1 : U2I= -2 : U3I= -.5 : U4I= 2 :

  UTIIA= (FVP*U1A + FFF*U2A + PFN*U3A + PVN*U4A) *24163;
  UTIIE= (FVP*U1E + FFF*U2E + PFN*U3E + PVN*U4E) *24163;
  UTIIC= (FVP*U1C + FFF*U2C + PFN*U3C + PVN*U4C) *24163;
  UTIID= (FVP*U1D + FFF*U2D + PFN*U3D + PVN*U4D) *24163;
  UTIIE= (FVP*U1E + FFF*U2E + PFN*U3E + PVN*U4E) *24163;
  UTIIF= (FVP*U1F + FFF*U2F + PFN*U3F + PVN*U4F) *24163;

PROC SORT DATA=DATA1 OUT=FIRST;
  BY DESCENDING UTIIA;
DATA FIRST;
  SET FIRST;
  KEEP CSCCRE PVP PFE PFN FVN SRATIO SUCCRATE UTIIA;
  IF N LE 30;
PROC PRINT;
  TITLE BASE UTILITY IS 19112 AND BASE SUCCESS RATE IS .8606;
  TITLE2;
  TITLE3 U1= 1 , U2= -.5 , U3= 0 , U4= .5 .;
PROC PICT DATA=FIRST;
  PICT UTIIA * SRATIC = '+' / VREF =19112;

PROC SORT DATA=DATA1 OUT=SECCND;
  BY DESCENDING UTIIE;
DATA SECCND;
  SET SECCND;
  KEEP CSCCRE PVP PFE PFN FVN SRATIO SUCCRA UTIIE;
  IF N LE 30;
PROC PRINT;
  TITLE BASE UTILITY IS 17428 AND BASE SUCCESS RATE IS .8606;
  TITLE2;
  TITLE3 U1= 1 , U2= -1 , U3= 0 , U4= 1 .;
PROC PICT DATA=SECCND;
  PICT UTIIE * SRATIC = '+' / VREF =17428;
```

```

PROC SORT DATA=DATA1 CUT=THIRD;
  BY DESCENDING UTIIC;
DATA THIRD;
  SET THIRD;
  KEEP CSCCRE PVP PFF PFN EVN SRATIO SUCCRATE UTIIC;
  IF N LE 30;
PROC PRINT;
  TITLE BASE UTILITY IS 14061 AND BASE SUCCESS RATE IS .8606;
  TITLE2;
  TITLE3 U1= 1 , U2= -2 , U3= 0 , U4= 2 .;
PROC FICT DATA=THIRD;
  PICT UTIIC * SRATIC = '+' / VREF =14061;

PROC SORT DATA=DATA1 CUT=FOURTH;
  BY DESCENDING UTIID;
DATA FOURTH;
  SET FOURTH;
  KEEP CSCCRE PVP PFF PFN EVN SRATIO SUCCRATE UTIID;
  IF N LE 30;
PROC PRINT;
  TITLE BASE UTILITY IS 19112 AND BASE SUCCESS RATE IS .8606;
  TITLE2;
  TITLE3 U1= 1 , U2= -.5 , U3= -.5 , U4= .5 .;
PROC FICT DATA=FOURTH;
  PICT UTIID * SRATIC = '+' / VREF =19112;

PROC SORT DATA=DATA1 CUT=FIFTH;
  BY DESCENDING UTIIF;
DATA FIFTH;
  SET FIFTH;
  KEEP CSCCRE PVP PFF PFN EVN SRATIO SUCCRATE UTIIF;
  IF N LE 30;
PROC PRINT;
  TITLE BASE UTILITY IS 17428 AND BASE SUCCESS RATE IS .8606;
  TITLE2;
  TITLE3 U1= 1 , U2= -1 , U3= -.5 , U4= 1 .;
PROC FICT DATA=FIFTH;
  PICT UTIIF * SRATIC = '+' / VREF =17428;

PROC SORT DATA=DATA1 CUT=SIXTH;
  BY DESCENDING UTIIF;
DATA SIXTH;
  SET SIXTH;
  KEEP CSCCRE PVP PFF PFN EVN SRATIO SUCCRATE UTIIF;
  IF N LE 30;
PROC PRINT;
  TITLE BASE UTILITY IS 14061 AND BASE SUCCESS RATE IS .8606;
  TITLE2;
  TITLE3 U1= 1 , U2= -2 , U3= -.5 , U4= 2 .;
PROC FICT DATA=SIXTH;
  PICT UTIIF * SRATIC = '+' / VREF =14061;

/*
//

```

LIST OF REFERENCES

1. Navy Enlisted Career Guide, 1980-81.
2. Center for Naval Analyses Report CAC 425, A New Look at Success Chances of Recruits Entering the Navy (SCREEN), by R. F. Lockman and P. M. Lurie, February 1980.
3. Center for Naval Analyses Report 81-0048, Relating Enlistment Standards to Job Performance: A Pilot Study for Two Navy Ratings, by P. M. Lurie, January 1981.
4. Center for Naval Analyses Report CRC 450, Continuous Estimates of Survival through Eight Years of Service Using FY 1979 Cross-Sectional Data, by P. M. Lurie, September 1979.
5. Center for Naval Analyses Report, Estimating Four-Year Survival and Reenlistment Probabilities for Boiler Technicians and Machinists' Mates, by J. W. Fletcher, February 1979.
6. Center for Naval Analyses Report, The Effect of Delayed Entry, Recruit Quality and Training Guarantees for Two-Year Survival of Navy Enlisted Men, by R. F. Lockman, 1976.
7. Center for Naval Analyses Report CRC 382, First Term Survival and Reenlistment Chances for Navy Ratings and a Strategy for their Use, by J. S. Thomas, May 1979.
8. Nesbitt, K. W., The Development of Selection Standards for Three Navy Ratings which Vary in Level of Complexity, MS Thesis, Naval Postgraduate School, Monterey, June 1983.
9. Snyder, W. I. and Bergazzi, W. A., Enlistment Standards for Two Navy Ratings: Boiler Technicians (BT) and Machinists' Mates (MM), MS Thesis, Naval Postgraduate School, Monterey, June 1983.
10. McGarvey, W. E., Application of Covariance Structure Analysis to Enlistment Standards, paper presented at the Industrial Management Society and the Operations Research Society of America joint meeting, San Francisco, California, 15 May 1984.
11. Berenson, M. I. and Levine, D. M., Basic Business Statistics, 2nd ed., Prentice Hall, Inc., 1983.

12. SAS Institute Inc., SAS User's Guide: Statistics, 1982 ed., SAS Institute Inc., 1982.
13. Campbell, J. F., Psychometric Theory, in Dunnette, M. D., Handbook of Industrial and Organizational Psychology, Rand McNally, 1976.
14. Guilford, J. F. and Fruchter, B., Fundamental Statistics in Psychology and Education, 5th ed., McGraw-Hill, 1973.
15. Navy Personnel Research and Development Center Report TR 82-37, Prediction of Job Performance: Review of Military Studies, by R. Vineberg and J. N. Jcyner, March 1982.
16. The Assessment Group Report R211, Billet Costs of Enlisted, Officer and Civilian Naval Personnel: FY 83, by O. L. Frankel and R. A. Butler, July 1983 (Updated March 1984).
17. Klecka, W. R., Discriminant Analysis, Sage Publications, Inc., 1980.
18. Nie, N. H., and others, Statistical Package for the Social Sciences, 2nd ed., McGraw-Hill, 1975.
19. Wiggins, J. S., Personality and Prediction: Principles of Personality Assessment, Addison-Wesley, 1973.

BIBLIOGRAPHY

Center for Naval Analyses Report 82-1357, Replacement Costs for Navy First Term Personnel by Rating, by E. Ballis and P. Clay-Bendéz, September 1982.

General Research Corporation Report CR-197, Development of Methods for Analysis of the Cost of Enlisted Attrition, by L. F. Huck and K. D. Eidlum, September 1977.

Sandel, C. D. and Gleason, M. F., Enlistment Standards as Related to Performance in Aviation Antisubmarine Warfare Operator and Aviation Antisubmarine Warfare Technician Ratings, MS Thesis, Naval Postgraduate School, Monterey, September 1983.

SAS Institute Inc., SAS User's Guide, 1982 ed., SAS Institute Inc., 1982.

The Assessment Group Report R207, The Billet Cost Model System, by A. A. Butler and O. L. Frankel, July 1983.

Whitwire, R. D. and Deitchman, C. G., An Enlisted Performance Prediction Model for Aviation Structural Mechanics, MS Thesis, Naval Postgraduate School, Monterey, August 1983.

INITIAL DISTRIBUTION LIST

	No. Copies
1. Defense Technical Information Center Cambridge Station Alexandria, Virginia 22314	2
2. Library, Code 0142 Naval Postgraduate School Monterey, California 93943	2
3. Department Chairman, Code 54 Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	1
4. Mr. Richard S. Elster Office of the Assistant Secretary of the Navy Room 4E778 The Pentagon Washington, DC 20350	1
5. Professor William E. McGarvey, Code 54Ms Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	2
6. Professor Thomas G. Swenson, Code 54Z Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	1
7. Professor George W. Thomas, Code 54Te Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	1
8. NITA Library, Code 36 Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	1
9. ICDE Dwayne A. Celund (CP-130F1) Office of the Chief of Naval Operations Navy Department Washington, DC 20350	2
10. CAPT J. S. A. Clark, RAE Personnel Branch-Army Fussell Offices Canberra, ACT 2600 Australia	2
11. Defence Library Campbell Park Offices Canberra, ACT 2600 Australia	2
12. Mr. Charles G. Deitchman (CP-111E) Office of the Chief of Naval Operations Navy Department Washington, DC 20350	1

13. Professor Ronald A. Weitzman, Code 54Wz
Department of Administrative Sciences
Naval Postgraduate School
Monterey, California 93943

1

END

FILMED

5-85

DTIC

